ABSTRACT

The correct simulation of midlatitude atmospheric blocking has always been a main concern since the earliest days of numerical modeling of Earth’s atmosphere. To this day blocking represents a considerable source of error for general circulation models from both a numerical weather prediction and a climate perspective. In the present work, 20 years of global climate model (GCM) developments are analyzed from the special point of view of Northern Hemisphere atmospheric blocking simulation. Making use of a series of equivalent metrics, three generations of GCMs are compared. This encompasses a total of 95 climate models, many of which are different—successive—versions of the same model. Results from model intercomparison projects AMIP1 (1992), CMIP3 (2007), and CMIP5 (2012) are taken into consideration. Although large improvements are seen over the Pacific Ocean, only minor advancements have been achieved over the Euro-Atlantic sector. Some of the most recent GCMs still exhibit the same negative bias as 20 years ago in this region, associated with large geopotential height systematic errors. Some individual models, nevertheless, have improved and do show good performances in both sectors. Negligible differences emerge among ocean-coupled or atmosphere-only simulations, suggesting weak relevance of sea surface temperature biases. Conversely, increased horizontal resolution seems to be able to alleviate the Euro-Atlantic blocking bias.

1. Introduction

Atmospheric blocking is a midlatitude weather pattern characterized by a quasi-stationary long-lasting, high pressure system that “blocks” and diverts the movement of the synoptic cyclones (Rex 1950). This usually occurs following the breaking of a Rossby wave at the exit of the storm track, when a subtropical low-vorticity air mass is advected poleward developing an anticyclonic circulation. Blocking occurs more frequently in winter and spring—when the jet stream is stronger—although it may be observed through the year.

Blocking has long elicited the interest of scientists. A slow large-scale Rossby wave, while able to create a large-scale pressure anomaly, cannot explain the persistence of a blocking event, which can remain in place for periods as long as several weeks. A number of theories involving nonlinear mechanisms have thus been proposed through the years. This includes solitonic solutions of the potential vorticity equation (McWilliams 1980; Malguzzi and Malanotte-Rizzoli 1984) or large-scale multiple equilibria of the atmospheric flow (Charney and DeVore 1979). The geographical position of blocking has been explained by the interaction of traveling Rossby waves with stationary eddies created by orography or land–sea contrast or with forced waves originating in the tropics (Egger 1978; Hoskins and Karoly 1981). Also, the interaction of transient baroclinic eddies with the large-scale blocking anticyclone has been shown to feed energy to the block and maintain it against dissipation (Green 1977; Shutts 1983; Nakamura et al. 1997).

Another more practical reason still makes blocking a critical topic: numerical models have been shown to have limited skill in reproducing it. In the early 1990s it was discovered that bad forecast of blocking accounted...
<table>
<thead>
<tr>
<th>Institute</th>
<th>AMIP1</th>
<th>AMIP3</th>
<th>CMIP3</th>
<th>AMIP5</th>
<th>CMIP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC, China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCC/RCC, Norway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BCC_CSM1.1</td>
</tr>
<tr>
<td>CCCma, Canada</td>
<td>CCC-GCM-II</td>
<td></td>
<td>BCC-BCM2.0</td>
<td>BCC_CSM1.1</td>
<td></td>
</tr>
<tr>
<td>CNRM-CERFACS, France</td>
<td>Météo-France–ARPEGE</td>
<td>CNRM-CM3</td>
<td>CNRM-CM3</td>
<td>CNRM-CM5</td>
<td>CNRM-CM5</td>
</tr>
<tr>
<td>COLA, United States</td>
<td>COLA-1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSIRO/CSIRO-BoM, Australia</td>
<td>CSIRO9 Mk1</td>
<td></td>
<td>CSIRO Mk3.0</td>
<td>ACCESS1.0</td>
<td>ACCESS1.0</td>
</tr>
<tr>
<td>CSU, United States</td>
<td>CSU-91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECMWF/EC-Earth consortium</td>
<td>ECMWF</td>
<td></td>
<td></td>
<td>EC-EARTH</td>
<td></td>
</tr>
<tr>
<td>Global Change and Earth System Science (GCESS), China</td>
<td></td>
<td></td>
<td></td>
<td>BNU-ESM</td>
<td>BNU-ESM</td>
</tr>
<tr>
<td>INGAV-CMCC, Italy</td>
<td></td>
<td></td>
<td>INGV-SXG</td>
<td>CMCC-CM</td>
<td>CMCC-ESM</td>
</tr>
<tr>
<td>INM, Russia</td>
<td></td>
<td>INM-CM3.0</td>
<td>INM-CM3.0</td>
<td>IPSL-CM5A-LR</td>
<td>IPSL-CM5A-LR</td>
</tr>
<tr>
<td>IPSL, France</td>
<td></td>
<td></td>
<td>IPSL-CM4</td>
<td>IPSL-CM5A-MR</td>
<td>IPSL-CM5A-MR</td>
</tr>
<tr>
<td>JMA, Japan</td>
<td>JMA-GSM</td>
<td></td>
<td>MIROC3.2 (medres)</td>
<td>MIROC5</td>
<td>MIROC5</td>
</tr>
<tr>
<td>MIROC, Japan</td>
<td></td>
<td>MIROC3.2 (hires)</td>
<td>MIROC3.2 (hires)</td>
<td>MIROC5</td>
<td>MIROC5</td>
</tr>
<tr>
<td>MIU, Germany and Korea</td>
<td></td>
<td>MIU-BECHOG</td>
<td>MIROC3.2 (medres)</td>
<td>MIROC5</td>
<td>MIROC5</td>
</tr>
<tr>
<td>MPI-M, Germany</td>
<td>MPI ECHAM3</td>
<td>MPI ECHAM5</td>
<td>MPI ECHAM5/OM</td>
<td>MPI-ESM-LR</td>
<td>MPI-ESM-LR</td>
</tr>
<tr>
<td>MRI, Japan</td>
<td>MRI-GCM-II</td>
<td>MRI-CGCM2.3.2</td>
<td>MRI-CGCM2.3.2</td>
<td>MRI-CGCM3</td>
<td>MRI-CGCM3</td>
</tr>
<tr>
<td>NASA, United States</td>
<td></td>
<td>GISS-ER</td>
<td>GISS-EH</td>
<td>CCSM4</td>
<td>CCSM4</td>
</tr>
<tr>
<td>NCAR, United States</td>
<td>NCAR CCM2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Meteorological Center (NMC), United States</td>
<td>NMC-MRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
for a large part of the systematic error of numerical weather prediction (Tibaldi and Molteni 1990; Tibaldi et al. 1994). This poor skill is still widely documented for general circulation models used for climate purposes (D’Andrea et al. 1998; Vial and Osborn 2012). Even the more recent analyses from the Coupled Model Intercomparison Project (Anstey et al. 2013; Masato et al. 2013; Dunn-Sigouin and Son 2013) still suggest the presence of a negative bias especially over Europe.

The origin of the underestimation of blocking frequency over Europe has been often connected with a wrong representation of the mean state that affects Rossby wave propagation (Scaife et al. 2010). However, it has been shown that good results can be obtained reducing the bias in the North Atlantic sea surface temperature (SST; Scaife et al. 2011) or increasing significantly the horizontal resolution of the atmospheric component of the models (Matsueda et al. 2009; Jung et al. 2012). Indeed, higher horizontal resolution is associated with improved transient eddy activity, which should sustain the blocking persistence (Berckmans et al. 2013), and with better resolved orography, which affects the mean state shaping the planetary waves (Brayshaw et al. 2009; Jung et al. 2012; Berckmans et al. 2013). Orography-related improvements can be obtained also with the introduction of parameterization that mimics the drag on the flow by subgrid orography (Jung et al. 2010; Pithan et al. 2016). Further reduction of blocking bias has been achieved with improved convection schemes, especially over the Pacific (Jung et al. 2010).

A number of objective definitions of blocking have been formulated for diagnostic studies. Blocking detection can be based on the gradient reversal of a specific dynamical field (Tibaldi and Molteni 1990; Pelly and Hoskins 2003), on high-amplitude anomalies of it (Dole and Gordon 1983), or on a combination of the two (Barriopedro et al. 2010). Geopotential height (Tibaldi and Molteni 1990), potential temperature (Pelly and Hoskins 2003), or vorticity (Schwierz et al. 2004) can be used in both monodimensional (D’Andrea et al. 1998) and bidimensional form (Scherrer et al. 2006; Tyrlis and Hoskins 2008).

The goal of this study is thus to assess the evolution—over the last 20 years—of the skill of global climate models in simulating Northern Hemisphere atmospheric blocking. To do so, we apply a series of equivalent metrics to three different subsets of a multimodel dataset. Data from the first Atmospheric Model Intercomparison Project (AMIP1, 1992), phase 3 of the Coupled Model Intercomparison Project (CMIP3, 2007), and phase 5 of the Coupled Model Intercomparison Project (CMIP5, 2012) are analyzed.

### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Institute</th>
<th>AMIP1</th>
<th>AMIP3</th>
<th>CMIP3</th>
<th>CMIP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA/GFDL, United States</td>
<td>GFDL</td>
<td>GFDL</td>
<td>GFDL</td>
<td>GFDL</td>
</tr>
<tr>
<td></td>
<td>CM1</td>
<td>CM3</td>
<td>CM2.1</td>
<td>CM2.1</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>GFDL-ESM2G</td>
<td>GFDL-ESM2M</td>
<td>GFDL-ESM2G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPN, Canada</td>
<td>NWP-D40P29</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Universities Global Atmospheric Modelling Programme (UGAMP), United Kingdom</td>
<td>UGCM-1.3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Met Office Hadley Centre (MOHC), United Kingdom</td>
<td>HadAM3</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

TABLE 1. (Continued)
D’Andrea et al. (1998) conducted a comprehensive analysis of the skill of models in reproducing blocking from AMIP1. They used a simple monodimensional definition of blocking based on the reversal of the meridional gradient of geopotential height, based on the Tibald and Molteni (1990) index. Even though possibly less dynamically significant than some of the more modern indices mentioned above, the same index as in D’Andrea et al. (1998) is chosen in this work. This is done in order to make the more modern models comparable with the earlier ones.

In section 2 below, the index is defined in detail and the datasets used are presented. Section 3 contains the main results, evaluating the models’ improvements in both blocking frequency and blocking duration. At the end of the section the relationship between blocking simulation and of the model systematic error is discussed, as well as the possible source of improvements, such as the horizontal resolution and the role of the SST. Section 4 contains a discussion and conclusions.

2. Data and blocking index

As mentioned in the introduction, data from the AMIP1 (Gates 1992; Gates et al. 1999), CMIP3 (Meehl et al. 2007), and CMIP5 (Taylor et al. 2012) are analyzed. When available, data from both coupled and atmosphere-only models are considered. For convenience, the ensemble of atmosphere-only models from CMIP3 and CMIP5 will be hereafter referred as AMIP3 and AMIP5. One ensemble member for each model has been examined, for a total of 95 different GCMs (15 from AMIP1, 9 from AMIP3, 20 from CMIP3, 22 from CMIP5, and 29 from CMIP5). The full list of models analyzed, grouped by simulation campaign and by modeling center, can be found in Table 1.

The ECMWF interim reanalysis (ERA-Interim) has been used as a reference (Dee et al. 2011). Blocking is studied on the common 10-yr period spanning from January 1979 to December 1988; this period is imposed by the original experimental protocol of the AMIP1. Although it can be considered too short, negligible differences have been found when the complete time windows for AMIP3, AMIP5, CMIP3, and CMIP5 datasets are considered (see Fig. S1 in the supplementary material).

The blocking index follows the definition of D’Andrea et al. (1998) but is adjusted in the different datasets in order to produce an equivalent metric. Before any computation, data are interpolated on a common 3.75° × 3.75° grid with a second-order conservative remapping method.

For ERA-Interim and AMIP5 and CMIP5 models, daily geopotential height at 500 hPa is used (Z500). We first define the following:

\[
GHGS(\lambda_0, \Delta) = \frac{Z500(\lambda_0, \phi_0 + \Delta) - Z500(\lambda_0, \phi_0 + \nabla)}{\phi_0 - \phi_5}
\]  

(1)

and

\[
GHGN(\lambda_0, \Delta) = \frac{Z500(\lambda_0, \phi_N + \Delta) - Z500(\lambda_0, \phi_0 + \Delta)}{\phi_N - \phi_0},
\]

(2)

where \(\Delta = -3.75^\circ, 0^\circ,\) and \(3.75^\circ; \phi_0 = 60^\circ\text{N}; \lambda_0\) ranges from \(0^\circ\) to \(360^\circ; \phi_5 = \phi_0 - 18.75^\circ;\) and \(\phi_N = \phi_0 + 18.75^\circ.\) Instantaneous blocking (IB) for a defined longitude is identified when at least for one of the three values of \(\Delta,\)

\[
IB(\lambda_0) = 1, \text{ if } \exists \text{ GHGS}(\lambda_0, \Delta) > 0 \land \text{ GHGN}(\lambda_0, \Delta) < -5 \text{ m (° lat)}^{-1}.
\]

(3)

For AMIP3 and CMIP3 models daily geopotential height is not available. Instantaneous blocking frequency is thus computed using the daily eastward wind at 500 hPa (U500), following the geostrophic relationship as done by Scaife et al. (2010). Indeed, the gradients in (1) and (2) can be expressed integrating the zonal wind in latitude:

\[
UGHGS(\lambda_0, \Delta) = \frac{2a\Omega}{g} \int_{\phi_5 + \Delta}^{\phi_0 + \Delta} U500(\lambda_0, \phi) \sin(\phi) \, d\phi
\]

(4)

and

\[
UGHGN(\lambda_0, \Delta) = \frac{2a\Omega}{g} \int_{\phi_0 + \Delta}^{\phi_N + \Delta} U500(\lambda_0, \phi) \sin(\phi) \, d\phi,
\]

(5)

where \(\Omega\) is Earth’s angular rotation velocity, \(a\) is Earth’s mean radius, and \(g\) is the gravitational acceleration. In this case, (3) becomes the following:

\[
IB(\lambda_0) = 1, \text{ if } \exists \text{ UGHGS}(\lambda_0, \Delta) < 0 \land \text{ UGHGN}(\lambda_0, \Delta) > 5(\phi_N - \phi_0).
\]

(6)

Conversely, original data for the AMIP1 dataset are no longer available. However, D’Andrea et al. (1996) used the same above-described geopotential height index to compute the blocking diagnostic in
the AMIP1 data and reported the blocking frequencies and duration for each model analyzed. Therefore, blocking data for AMIP1 experiments were graphically retrieved from D’Andrea et al. (1996); this has been possible making use of WebPlotDigitalizer (http://arohatgi.info/WebPlotDigitizer/), a free online tool.

The equivalence of the three metrics—geopotential height index, zonal wind index, and graphically retrieved geopotential height index from D’Andrea et al. (1996)—can be verified using ERA-Interim. The instantaneous blocking climatology—expressed as blocking frequency (i.e., percentage of blocked days)—for DJF 1979–88 is shown in Fig. 1, where the black (geopotential height index) and blue (zonal wind index) lines are indeed almost indistinguishable. Figure 1 also shows in red the graphically retrieved metric computed from the ECMWF operational analysis for the AMIP1 period (DJF 1979–88). It can be seen that this curve also agrees with the others; the error for blocking frequency (measured as the weighted average of the relative difference on each grid point) is ±5%, even considering that the ECMWF operational analysis is slightly different from ERA-Interim. The error associated with the graphic retrieval tool alone has been estimated to be ±2%.

The equivalence of the zonal wind and geopotential height metric is robust not only under a climatological point of view but also from a daily perspective; the correlation between the two daily indices is 0.88, computed over the 903 winter days × 96 longitudes (86,688 grid points).

In addition to the blocking frequencies, we also analyze blocking duration. To this aim, we introduce the definition of “blocking event” as opposed to the above-defined “instantaneous blocking”; we use the same sector definitions and spatiotemporal constraints as D’Andrea et al. (1998).

Euro-Atlantic blocking is evaluated for 26.25°W–41.25°E, whereas Pacific blocking is evaluated for 116.25°E–146.25°W. A sector is considered as blocked in a certain day if at least three adjacent longitudes (11.25°) are blocked. To satisfy the synoptic definition of block (Rex 1950)—and thus in order to define a blocking event—a time persistence of minimum five consecutive days is applied. Two exceptions are possible: 1) if two blocked successive days are followed by one nonblocked day and then other two blocked days, the nonblocked day is considered as blocked; and 2) if a single nonblocked day is followed (preceded) by four blocked days and it is preceded (followed) by a single blocked day, the nonblocked day is considered as blocked.

The resulting binary time series define the occurrence of blocking events, and they allow the computation of the duration of each event over both the Euro-Atlantic and Pacific sectors. Please note that blocking events are characterized by a blocking event frequency, which is expressed as the percentage of
3. Results

a. Blocking frequency

As mentioned in the introduction, GCMs are known to underestimate blocking frequency, over both the Pacific and Euro-Atlantic sector (e.g., Amstey et al. 2013). In Fig. 2, the winter (DJF) instantaneous blocking frequencies from nine different modeling centers whose models have being part of both the AMIP1 and AMIP5 simulation campaign are shown. About 20 years divide the results plotted in green and those in blue. Yellow lines show, when available, the results from the corresponding AMIP3 models. Putting the output of different models together on the same graph is instructive but should be done with caution. Some of the modeling centers use evolution of the same atmospheric models (e.g., MPI, GFDL, and MRI), while others have been subject to radical changes (e.g., CSIRO and CCCma). Therefore, when some form of lineage between two models could be established (see Table 1 or also Knutti et al. 2013), they were included in Fig. 2. When more than one model is available from a certain simulation campaign the one with higher horizontal resolution is retained.

The emerging scenario shows the presence of a different level of improvements among models. First, the overestimation of blocking over the western Atlantic (broadly corresponding to the high-latitude blocking over Greenland; Davini et al. 2012) that was present in a few old generation models (HadAM1, CSIRO9 Mk1, and MRI-CGCM2) has been resolved. Second, the large negative bias over the Pacific has been corrected by many models (above all MPI-ESM-MR and MRI-CGCM3). However, less agreement is seen over the Euro-Atlantic sector, well known for being poorly represented by both CMIP3 and CMIP5 models (Scaife et al. 2010; Amstey et al. 2013). None of these AMIP5 models is able to reach the ERA-Interim frequency, and some of them only show weak improvements between the AMIP1 and AMIP5 version, or none at all (NCAR and CNRM–CERFACS models).

The different intercomparison projects can be considered together analyzing the multimodel mean (MMM) of each group. Figure 3 shows MMM for the
DJF instantaneous blocking frequencies for AMIP1, AMIP3, AMIP5, CMIP3, and CMIP5. The number of models analyzed within each cluster is shown in the legend. At a glance, whereas large improvements are evident over the Pacific sector, minor advancements have been obtained over the Atlantic quadrant.

To assess the statistical significance of the differences in Fig. 3, we performed the Welch $t$ test with a 5% significance level. We compared the MMM of different pairs of simulation campaigns in the Euro-Atlantic and Pacific regions, defined by integrating the curves in Fig. 3 between 26.25°W–41.25°E and 116.25°E–146.25°W, respectively. Over the Euro-Atlantic sector no statistical significant differences are found—given the large spread in blocking frequency between the GCMs belonging to each intercomparison project. Conversely, over the Pacific AMIP5 and AMIP3 significantly outperform AMIP1 (even if no difference is found between AMIP5 and AMIP3) and CMIP5 is found to do better than CMIP3.

Outside these two main regions, with the exclusion of the anomalous results of the AMIP3 models, no statistically significant improvements are found over the western Atlantic–Greenland region (75°–37.5°W). Conversely, CMIP5 and AMIP5 seem to do better than their predecessors in the Ural blocking region (30°–90°E). It is also important to highlight that even though the considered time window is rather short (10 years), the results from the MMM analysis are definitely robust. Indeed, Fig. S1 in the supplementary material shows that negligible differences emerge when the largest available time window of each dataset is used.

To summarize, four main considerations can be drawn from Fig. 3:

1) Blocking frequency over the Pacific and the Ural regions is much better represented in the recent generation of GCMs, with the MMM of CMIP5 models showing a minor negative bias.
2) Negligible improvements are obtained over the Euro-Atlantic and western Atlantic (i.e., Greenland) sectors.
3) Minor differences are seen between coupled and atmosphere-only experiments, suggesting weak relevance of SSTs.
4) Overall, the differences between CMIP3/AMIP3 and CMIP5/AMIP5 are small (with the exception of CMIP5 outperforming CMIP3 over the Pacific sector), confirming that CMIP5 should be considered as an improved ensemble of the same models rather than an ensemble of new climate models (Knutti et al. 2013).

Figure 4 shows the distribution of the climate models of each modeling campaign according to their Euro-Atlantic and Pacific instantaneous blocking frequency bias. It can be seen that, even if the MMM is slightly affected, in the more recent modeling campaigns the number of climate models with a small bias is increasing considerably (see the percentages plotted at the top left of both panels in Fig. 4). For instance, about the 31% (52%) of CMIP5 models have a bias that is...
smaller than or equal to the 20% over the Euro-Atlantic (Pacific), much better than any of their predecessors. However, Fig. 4 shows also that over the Pacific an increasing number of models is producing an excess of blocking.

Overall this is encouraging, suggesting that some modeling centers are now able to resolve atmospheric blocking in a reasonable way. Nonetheless, it must be kept in mind that even if for CMIP5 there has been a large increase of the number of models available (22 for AMIP5 and 29 for CMIP5), many of them share significant portions of their codes or even use the same atmospheric model. For instance, it should be remarked that within the nine models with a minor bias over the Euro-Atlantic (i.e., with a bias $\leq 20\%$) five run the ECHAM5 atmospheric model (CMCC-CM, CMCC-CMS, MPI-ESM-LR, MPI-ESM-MR, and MPI-ESM-P). The other four models within this group are HadGEM2-CC and ACCESS1.0 (both running a version of the atmospheric component of the Met Office Unified Model), EC-EARTH, and MIROC5.

b. Blocking event duration

The introduction of the spatial and temporal constraints presented in section 2 leads to the definition of blocking event duration over both the Euro-Atlantic and Pacific sectors. As done in D’Andrea et al. (1996, 1998), the analysis of duration involves the full year and not only the winter season.

D’Andrea et al. (1998) described the exponential shape of the distribution of blocking duration and explained it as a sign of the fact that blocking decay is a memoryless process, similar to a random walk. In other words, the probability of blocking of surviving from day $n$ to day $n + 1$ is not depending on $n$. The exponential distribution of duration has been observed also in more recent coupled GCMs (e.g., Vial and Osborn 2012).
To be more precise, one should however speak of geometric distribution (i.e., the discrete form of the exponential). This is also a one-parameter distribution in which the mean $m$ and standard deviation $s$ are linked by the relation $s = \left[ \frac{1}{m} \right]^{1/2}$. In our case, because of the 5-day minimum requirement for duration, the relation between mean and standard deviation is slightly shifted and becomes exactly equal to $s = \left[ \frac{1}{m} \right]^{1/2}$ (see the discussion in the appendix). This is confirmed in Fig. 5 where all the models approximately follow the theoretical ratio between mean and standard deviation (shown by the red line). Although Masato et al. (2009) showed a deviation from this simple model for long-lasting events, the geometric distribution remains a reasonable approximation to describe the distribution of blocking duration.

Figure 5 also shows that CMIP5 and AMIP5 models have longer average duration than their older counterparts (i.e., CMIP3, AMIP3, and AMIP1). However, these are still lower than ERA-Interim, with the exception of CMIP5 over the Pacific. This can be explained looking at the blocking events distribution; when compared to ERA-Interim, models tend to underestimate the number of long-lasting events in favor of higher counts of short-lived events (not shown), as reported also by Vial and Osborn (2012).

To have a robust sample to analyze, we grouped together the blocking events from all the models from each intercomparison project, computing an average number of events, an average duration, and an average number of model days for each of those. These are shown in the left panel of Fig. 6; note that these results account for full-year blocking events.

The underestimation of blocking events is again evident over the Euro-Atlantic, especially in terms of duration. Over the Pacific, good results are achieved by CMIP5 models.

But what does govern the relationship between the bias in blocking duration and blocking frequency? A model can “increase” its blocking frequency in two ways: increasing the number of blocking events (i.e., with more blocking onsets—for instance, with a larger number of Rossby waves breaking at the exit of the jet) and/or increasing the duration of the events (i.e., increasing the blocking persistence with a more effective maintenance mechanism). While the former will affect only the number of counts, the latter will change the shape of the geometric distribution, with a relative increase of long-lasting events.

The blocking event frequency (i.e., the number of blocked days that fits the blocking event definition) can be defined by a simple formula $\text{Freq} = \frac{m N_{\text{events}}}{N_{\text{days}}}$, where $m$ is the mean duration, $N_{\text{events}}$ the number of events, and $N_{\text{days}}$ the number of model days. The relative contribution of increased duration or increased number of events to a change of frequency can be obtained by finite-differentiating the previous definition:

$$
\Delta \text{Freq} = \frac{N_{\text{events}} \Delta m}{N_{\text{days}}} + \frac{\mu N_{\text{events}}}{N_{\text{days}}} \Delta N_{\text{events}}. 
$$
In this equation, the first term on the rhs represents the contribution to the change of frequency due to the increased "maintenance" of blocking, and the second term represents that due to the increased number of "onset."

We apply this decomposition comparing the more recent clusters (CMIP5 and AMIP5) to their older counterparts (CMIP3, AMIP3, and AMIP1). Results are shown in the right panel of Fig. 6. The largest contribution to the higher observed blocking frequencies is given by the onset term (i.e., larger number of blocking events), over both the Pacific and Euro-Atlantic regions. For the AMIP models this predominance is very strong (about 80%–90% of the increase in frequency can be explained by the onset term), while for CMIP it is less evident but still remarkable (about 65% due to onset and 35% due to maintenance).

These results suggest a possible distinction between the mechanisms for onset and maintenance of blocking events. Considering also that coupled models—especially in the more recent model intercomparison project—have longer average duration (as shown by Fig. 6), it is intriguing to consider the possibility that the dynamical feedback due to air–sea interactions could potentially increase the block lifetime.

c. Sources of blocking bias

1) SST

As seen in Fig. 3, negligible differences are found between AMIP and CMIP (i.e., among atmosphere-only and coupled simulations). This suggests a weak influence of SST on winter instantaneous blocking frequencies. To further test this hypothesis, we considered the models from CMIP3 and CMIP5 that have both a coupled and atmosphere-only version available. The number of models analyzed thus decreases to 9 for CMIP3 and to 18 for CMIP5. Even within this specific subset, over the Euro-Atlantic the differences between coupled and atmosphere-only simulations are still negligible (Fig. S3 in the supplementary material).

Going into further detail, we considered each model separately to provide a fair comparison where only the SST bias and the air–sea coupling take part; we thus evaluate the bias reduction in instantaneous blocking frequencies of the atmosphere-only model (i.e., AMIP) with respect to its coupled counterpart (i.e., CMIP). Finally we regressed such changes in blocking frequencies against the root-mean-square error of the Atlantic and Pacific SSTs of each coupled model, using the HadISST dataset (Rayner et al. 2003) as reference (not shown).

With the exception of a few models showing a very large SST bias that improve considerably in their AMIP version (e.g., the CMIP3 model FGOALS-g1.0 and the CMIP5 IPSL model family, where the SST bias is on the order of several degrees Celsius), none of the above-mentioned tests produced a positive result that could link SST biases and blocking frequencies. Actually about the half of the models deteriorate their
blocking frequencies when used in AMIP mode, suggesting that blocking may benefit from dynamically active air–sea interactions, leading to better blocking frequencies in coupled GCMs than in atmosphere-only ones.

This seems to be especially true for the Pacific sector, where CMIP5 models outperform AMIP5 blocking frequency by an average relative increase of 20.2%; this can be seen also in Fig. 3. The sensitivity of Pacific blocking to oceanic boundary conditions was also observed by Tibaldi et al. (1997) in the AMIP1 version of the ECHAM model. However, this difference is rather small and does not pass a significance test at the 5% level (with the Welch t test). The same analysis has been performed on the largest possible common period of AMIP5 and CMIP5 (DJF 1979–2004). Here the difference remains insignificant and is even smaller, pointing to the absence of a robust distinction between the datasets.

It has been argued that the simulation of Atlantic jet stream and blocking variability may be affected by the North Atlantic SST errors and by the properties of the midlatitude SST frontal zone (Sampe et al. 2010; Keeley et al. 2012; O’Reilly et al. 2015). Furthermore, Scaife et al. (2011) found that a reduced SST bias considerably improves blocking climatology. However, our findings suggest that—in the Euro-Atlantic as in the Pacific—there are no significant differences in winter atmospheric blocking between coupled and atmosphere-only models. In this sense, our results are in agreement with other recent works with CMIP5 models on blocking and storm track (Masato et al. 2013; Zappa et al. 2014).

We can see two different explanations for those contradicting results: On the one side, it is possible that the benefits of a fully resolved midlatitude SST front as in the case of atmosphere-only simulation can be exploited only if a sufficiently high horizontal resolution is used (O’Reilly et al. 2015), which is not the case for the AMIP3 and AMIP5 models. On the other side, each GCM is known for having different levels of sensitivity to SST patterns, and therefore specific model-dependent features could be hidden by the multimodel comparison here performed.

2) HORIZONTAL RESOLUTION

Increasing horizontal resolution is the most invoked explanation for the long-lasting problem of atmospheric blocking underestimation, especially in the Euro-Atlantic quadrant. In Fig. 7 this relationship is explored, using the DJF 1979–88 instantaneous blocking frequency averaged over the Euro-Atlantic and Pacific sectors as a metric. The analyzed models span from more than 5° to less than 1° of longitudinal resolution (at the equator). The left panel of Fig. 7 shows that there is a statistically significant correlation (99% confidence level) between horizontal resolution and blocking bias over Europe, which suggests that increased resolution favors the reduction of the bias.
The importance of a correct representation of transient eddy activity and a better-resolved orography—both consequences of an increased horizontal resolution—has often been thought to be relevant for Euro-Atlantic blocking (Jung et al. 2012; Berckmans et al. 2013); thus it is not surprising that following a resolution increase the Euro-Atlantic blocking bias is reduced.

Conversely, over the Pacific the correlation between model bias and resolution is not statistically significant. The reduction of the bias is here instead associated with the more recent generation of GCMs (with AMIP5/CMIP5 outperforming AMIP3/CMIP3). Given the well-known relationship connecting Pacific blocking and tropical variability (e.g., Renwick and Wallace 1996; Carrera et al. 2004), it is not unlikely that the observed reduced bias over the Pacific is more connected to improved model physics rather than simply higher horizontal resolution. Indeed, Jung et al. (2010) found that for the

![Figure 8](image)
The implementation of an improved convection scheme leads to reduction of the DJF Pacific blocking bias.

### 3) Systematic Errors and the Role of the Mean Flow

The CMIP5 models’ Z500 still show large systematic errors, characterized by flow being too zonal (e.g., Anstey et al. 2013). In Fig. 8 the winter Z500 systematic error for the full AMIP5 winter (DJF) period 1979–2008 is shown for the nine models part of both AMIP1 and AMIP5. If compared to Fig. 9 from D’Andrea et al. (1998), it is possible to see that the AMIP5 models present geopotential height errors with the same order of magnitude of the old AMIP1 models.

A question that is still open is how much blocking bias is the cause of the systematic error and how much is caused by it (Tibaldi and Molteni 1990). Considering a given phenomenon, blocking in this case, with an observed frequency \( f \) and a simulated frequency \( f' \), we can decompose the systematic Z500 error as follows:

\[
z - \hat{z} = (f - \hat{f})(z_b - \hat{z}_b) + \hat{f}(z_b - \hat{z}_b - (z_c - \hat{z}_c) + (z_c - \hat{z}_c).
\]

The first term on the rhs of (9) indicates the part of the systematic error due to a bias in the frequency of blocking, the second represents the error related to the misrepresented pattern of blocking anomaly, and the last term indicates the systematic error when no blocking is occurring. The linear combination of the blocking frequency and the blocking pattern errors is a measure of the systematic error associated with blocking.

We applied this technique to the DJF of AMIP5 models on the 1979–2008 period using the blocking event frequency as a discriminant; the results, measured with the root-mean-square error (RMSE) over the 30°–75°N, 26.25°W–41.25°E region, are reported in Table 2. Figure 9 shows as an example the error decomposition for ACCESS1.0—that is, the model for which the ratio of blocking/nonblocked error is largest (see last column of Table 2). Conversely, Fig. 10 shows the decomposition for the model for

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Climatological</th>
<th>Frequency</th>
<th>Pattern</th>
<th>Blocked</th>
<th>Nonblocked</th>
<th>Blocked/nonblocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC-EARTH</td>
<td>31.84</td>
<td>7.97</td>
<td>7.06</td>
<td>10.34</td>
<td>26.01</td>
<td>0.4</td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>20.06</td>
<td>12.04</td>
<td>8.88</td>
<td>12.63</td>
<td>17.6</td>
<td>0.72</td>
</tr>
<tr>
<td>CCSM4</td>
<td>42.81</td>
<td>22.04</td>
<td>13.38</td>
<td>16.72</td>
<td>29.97</td>
<td>0.56</td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>33.51</td>
<td>16.21</td>
<td>9.67</td>
<td>17.22</td>
<td>20.6</td>
<td>0.84</td>
</tr>
<tr>
<td>GFDL CM3</td>
<td>42.56</td>
<td>19.71</td>
<td>9.87</td>
<td>17.7</td>
<td>32.74</td>
<td>0.54</td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>29.42</td>
<td>18.5</td>
<td>5.92</td>
<td>21.89</td>
<td>24.96</td>
<td>0.88</td>
</tr>
<tr>
<td>HadGEM2-A</td>
<td>19.55</td>
<td>4.29</td>
<td>10.65</td>
<td>10.18</td>
<td>16.48</td>
<td>0.62</td>
</tr>
<tr>
<td>CanAM4</td>
<td>54.85</td>
<td>20.99</td>
<td>8.23</td>
<td>23.05</td>
<td>40.02</td>
<td>0.58</td>
</tr>
<tr>
<td>ACCESS1.0</td>
<td>22.2</td>
<td>13.19</td>
<td>11.06</td>
<td>18.34</td>
<td>14.52</td>
<td>1.26</td>
</tr>
<tr>
<td>ACCESS1.3</td>
<td>26.93</td>
<td>14.03</td>
<td>12.35</td>
<td>16.71</td>
<td>16.88</td>
<td>0.99</td>
</tr>
<tr>
<td>FGOALS-g2</td>
<td>48.4</td>
<td>15.41</td>
<td>13.36</td>
<td>27.38</td>
<td>56.09</td>
<td>0.49</td>
</tr>
<tr>
<td>FGOALS-s2</td>
<td>48.07</td>
<td>21.79</td>
<td>13.45</td>
<td>23.3</td>
<td>34.4</td>
<td>0.68</td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>31.32</td>
<td>12.26</td>
<td>7.11</td>
<td>13.77</td>
<td>25.9</td>
<td>0.53</td>
</tr>
<tr>
<td>BCC_CSM1.1</td>
<td>62.59</td>
<td>18.51</td>
<td>12.8</td>
<td>23.09</td>
<td>52.68</td>
<td>0.44</td>
</tr>
<tr>
<td>BCC_CSM1.1(M)</td>
<td>59.78</td>
<td>20.41</td>
<td>15.75</td>
<td>16.35</td>
<td>49.13</td>
<td>0.33</td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>51.41</td>
<td>23.08</td>
<td>10.61</td>
<td>27.19</td>
<td>44.5</td>
<td>0.61</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>46.79</td>
<td>19.23</td>
<td>10.83</td>
<td>20.38</td>
<td>43.44</td>
<td>0.47</td>
</tr>
<tr>
<td>IPSL-CM5B-LR</td>
<td>72.24</td>
<td>14.29</td>
<td>7.42</td>
<td>20.14</td>
<td>70.76</td>
<td>0.28</td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>43.35</td>
<td>27.4</td>
<td>17.56</td>
<td>21.24</td>
<td>29.18</td>
<td>0.73</td>
</tr>
<tr>
<td>MIROC5</td>
<td>38.43</td>
<td>7.06</td>
<td>8.86</td>
<td>14.88</td>
<td>48.42</td>
<td>0.31</td>
</tr>
<tr>
<td>BNU-ESM</td>
<td>46.87</td>
<td>22</td>
<td>10.33</td>
<td>25.44</td>
<td>41.7</td>
<td>0.61</td>
</tr>
<tr>
<td>Mean</td>
<td>41.57</td>
<td>16.69</td>
<td>10.72</td>
<td>18.95</td>
<td>35.05</td>
<td>0.61</td>
</tr>
</tbody>
</table>

ECMWF model the implementation of an improved convection scheme leads to reduction of the DJF Pacific blocking bias.

#### Table 2. RMSE over the Euro-Atlantic region (30°–75°N, 26.25°W–41.25°E) following systematic error decomposition according to blocking event frequency over the DJF 1979–2008 period. Because of the nonlinearity in signs the sum of frequency, pattern, and non-blocked values do not equal climatological value. Units are meters. Last column shows the ratio between the blocked condition (combination of frequency and pattern) and the nonblocked flow.
which the bias is more associated with nonblocked flow (EC-EARTH).

In general, the error related to the blocking pattern is limited and in 19 of 21 models is smaller than the error associated with the blocking frequency. This suggests that although AMIP5 models underestimate the Euro-Atlantic occurrence, when this occurs its simulation is well performed. Moreover, the amount of error associated with the blocking frequency is highly variable among models, suggesting that the causes related to blocking negative bias over Europe can be numerous and different in different models.

On average, the error associated with blocking is 61% of the error associated with the nonblocked flow—and only in one case the former is larger than the latter—suggesting that the lack of blocking is only a part of the story. It is thus possible to state that the largest source of error is still placed in the simulation of the “zonal flow” for almost the totality of the models.

The same decomposition has been applied for the Pacific sector (see Table 3); as expected, considering the reduced blocking bias in this region, the error associated with the blocking is less important (about 46% of the error associated the nonblocked flow). In more detail, the error associated with blocking frequency is considerably smaller, with only 4 of 21 models for which this error is larger than the error associated with the blocking pattern.

As a final note, it should be made clear that this decomposition is mainly diagnostic; it extracts three components from the systematic error pattern. The three components are in no way independent from a dynamical point of view; for instance, a reduction of the nonblocked error—which is associated with the climatological position and strength of the eddy-driven jet—would certainly...
also influence the frequency and the pattern of blocking, thus likely projecting onto the two other components.

4. Discussion and conclusions

In this work we examined the evolution of atmospheric blocking modeling in the last 20 years, from the AMIP1 experiments to the more recent CMIP5. We made use of a series of equivalent metrics to investigate the blocking frequency and duration, with special regard to the Euro-Atlantic and Pacific sectors. Almost 100 models have been analyzed over the common time window of 1979–88, and larger periods have been used when possible to strengthen the robustness of the conclusions.

Results show clear improvements for winter Pacific blocking, specifically from AMIP1 to AMIP3 and AMIP5 and from CMIP3 to CMIP5. On the contrary, minor and not significant changes are observed for winter Euro-Atlantic blocking. However, a few models are considerably successful in reducing their bias in their most recent version over both sectors.

Blocking event duration for the full year has been also investigated; its distribution is consistent with a memoryless process giving rise to a geometric distribution. On average, more recent models simulate blocking with longer lifetimes—although an underestimation with respect to ERA-Interim is still evident, especially over the Euro-Atlantic. Increased blocking frequency is associated with both a larger number of blocking events (i.e., more onsets) and longer blocking lifetimes (i.e., more effective maintenance mechanisms). When the two contributions are compared, the onset term is found to be more important than the maintenance one, especially for atmosphere-only models.

An increase in horizontal resolution seems to be important to improve the representation of wintertime blocking over Euro-Atlantic. On the contrary, benefits from higher resolution are not evident in the Pacific sector. The phenomenological and dynamical differences among the two types of blocking events (Davini et al. 2012) are consistent with this different behavior: Pacific blocking is in fact more affected by tropical dynamics [e.g., El Niño–Southern Oscillation (see Carrera et al. 2004) or more generally Rossby
wave packets generated by tropical precipitation]. Also, no significant changes are observed between atmosphere-only and coupled models, suggesting a marginal relevance of the SST bias. In any case, specific model dependency has been found, observing that some models positively react to improved SSTs while others even deteriorate their blocking climatology.

A comment on the possible source of the wintertime geopotential height systematic error in the AMIP5 models has been presented; although the systematic errors of the models are generally quantitatively comparable, the dynamical explanation behind such bias appears to be different. It is found that the nonblocked flow (i.e., when no blocking is occurring and the flow is mainly zonal) is the most relevant source of error for many models. Over the Euro-Atlantic, blocking pattern is generally well represented by the models, whereas blocking frequency can contribute significantly to the total systematic error of the model. A reverse situation, with larger values for blocking pattern error, is observed over the Pacific.

Such different results from the systematic error decomposition suggest a final note of caution on the usage of multimodel means. If on the one side multimodel comparison can provide useful projections for climate scenarios—providing to the community a common “consensus” for future climate—on the other side they can create confusion when the analysis is aimed at identifying the processes at the origin the biases. We indeed must keep in mind that MMM gives an equal weight to all the models, which often share significant portion of the code, and that at the same time it is unduly unfair to equally consider GCMs from different modeling centers that have completely different budget and manpower availability.

We thus conclude by highlighting that although the overall picture seems somewhat discouraging—with good improvements over the Pacific but negligible advancements over Euro-Atlantic blocking in 20 years—the modeling community appears to be moving on the right track, with slow but constant improvements (see left panel of Fig. 6), for at least two reasons: On the one hand the continuously increasing availability of computational power is making possible climate runs of increasing resolution, which will most likely affect blocking representation, especially over the Euro-Atlantic sector. On the other hand, improvement in dynamics and new parameterization schemes—as orographic drag, convection, or stochastic physics schemes—have been shown to contribute to improve the representation of the midlatitude climate variability (e.g., Jung et al. 2010; Berner et al. 2012). Further sources of improvement may come from the representation of the stratosphere, sometimes associated with the refinement of the vertical grid (Anstey et al. 2013). Therefore, it is likely that the problem of blocking simulation in GCMs will be significantly alleviated in future years.

Acknowledgments. The authors want to thank all modeling groups participating in CMIP3 and CMIP5 for producing and sharing model data. We gratefully acknowledge the funding from the European Union’s Horizon 2020 Research and Innovation Programme COGNAC under the
European Union Marie Skłodowska-Curie Grant Agreement 654942. The authors would like to thank Stefano Tibaldi and three anonymous reviewers for their constructive comments that helped to improve this manuscript.

APPENDIX

Mean and Variance of a Geometric Distribution of Blocking Duration

Suppose that the probability of blocking decay at a given day is $\alpha$. Then, if there are $N_0$ blocks at time 0, there will be $(1 - \alpha)N_0$ blocks at time 1, $(1 - \alpha)^2 N_0$ at time 2, and so on. In synthesis, given this assumption, the probability distribution of the block duration is given by

$$p(N) = \alpha (1 - \alpha)^N. \quad (A1)$$

This is a typical geometric distribution, which has mean $\mu = (1 - \alpha)/\alpha$ and standard deviation $\sigma = [(1 - \alpha)/\alpha^2]^{1/2}$ [see, e.g., Gut (2012) or any probability textbook]. Eliminating $\alpha$ in these two last formulas gives the relation between $\mu$ and $\sigma$ given in the text. In our case, though, $N$ ranges from 5 to 4, and the mean and variance are hence given by the following:

$$\mu = \sum_{N=5}^{4} N \frac{\alpha (1 - \alpha)^N}{1 - S} \quad (A2)$$

$$\sigma = \sum_{N=5}^{4} (N - \mu) \frac{\alpha (1 - \alpha)^N}{1 - S}, \quad (A3)$$

where $(1 - S)$ is a normalization constant. In fact, since the distribution in (A1) is normalized, then $\sum_{N=5}^{4} p(N) = (1 - S)$, where $S$ is given by the limited sum $\sum_{N=5}^{4} \alpha (1 - \alpha)^N$.

The sums in (A2) and (A3), as well as the normalization constant $S$, can be computed analytically and give the following:

$$\left\{ \begin{align*}
\mu &= \frac{1 + 4\alpha}{\alpha} \\
\sigma &= \sqrt{\frac{(1 - \alpha)}{\alpha^2}}
\end{align*} \right.$$

Eliminating $\alpha$ in the above linear system gives the relation $\sigma = [(\mu - 4)/(\mu - 5)]^{1/2}$, which is exactly the red line plotted in the two panels of Fig. 5.

REFERENCES


