Climate Dynamics manuscript No.
(will be inserted by the editor)

Current and future atmospheric circulation at 500 hPa over Greenland simulated by the CMIP3 and CMIP5 global models

Alexandre Belleflamme · Xavier Fettweis · Charlotte Lang · Michel Erpicum

Received: date / Accepted: date

Abstract The Greenland ice sheet is projected to be strongly affected by global warming. These projections are either issued from downscaling methods (such as Regional Climate Models) or they come directly from General Circulation Models (GCMs). In this context, it is necessary to evaluate the accuracy of the daily atmospheric circulation simulated by the GCMs, since it is used as forcing for downscaling methods. Thus, we use an automatic circulation type classification based on two indices (Euclidean distance and Spearman rank correlation using the daily 500 hPa geopotential height) to evaluate the ability of the GCMs from both CMIP3 and CMIP5 databases to simulate the main circulation types over Greenland during summer. For each circulation type, the GCMs are compared to three reanalysis datasets on the basis of their frequency and persistence differences. For the current climate (1961-1990), we show that most of the GCMs do not reproduce the expected frequency and the persistence of the circulation types and that they simulate poorly the observed daily variability of the general circulation. Only a few GCMs can be used as reliable forcings

A. Belleflamme
Laboratory of Climatology and Topoclimatology
University of Liège
Allée du 6 Aout, 2, 4000 Liège, Belgium
Tel.: +32-43-665354
Fax: +32-43-665722
E-mail: A.Belleflamme@ulg.ac.be

X. Fettweis, C. Lang and M. Erpicum
Laboratory of Climatology and Topoclimatology, University of Liège, Belgium

for downscaling methods over Greenland. Finally, when applying the same approach to the future projections of the GCMs, no significant change in the atmospheric circulation over Greenland is detected, besides a generalised increase of the geopotential height due to a uniform warming of the atmosphere.

Keywords General Circulation Models · 500 hPa geopotential height · Greenland · Circulation Type Classification

1 Introduction

Atmosphere-Ocean General Circulation Models (AOGCMs) project that, in the future, global warming will be much more important in the polar regions and particularly in the Arctic compared to other regions ([Meehl et al. 2007]). Moreover, recent observations show that the Greenland ice sheet (GrIS) climate is warming and that a part of these changes are attributable to the general circulation ([Hanna et al. 2008], [Hanna et al. 2009], [Tedesco et al. 2008], [Box et al. 2010], [Fettweis et al. 2011b], [Box et al. 2012] and [Fettweis et al. 2012]). Indeed, [Fettweis et al. 2011b], [Mote 1998a] and [Mote 1998b] showed that there is a strong link between atmospheric circulation and near-surface air temperature (impacting the surface snow melt) over the Greenland ice sheet. [Mote 1998a] analysed teleconnections and [Mote 1998b] performed a cluster analysis; both analyses were based on a principal component analysis to study the linkage between circulation patterns at 700 hPa over the whole Arctic region and the Greenland ice sheet melt. They showed that the melting rate can be very different from one circulation pattern to another and that a significant part of the current trend towards increasing melt can be explained by changes in the atmospheric circulation.

In addition, it is known that General Circulation Models (GCMs) better simulate the general circulation than surface variables such as temperature or precipitation ([Yarnal et al. 2001]). Indeed, the coarse resolution of GCMs makes it very difficult to reliably simulate surface variables, which have important local variations and are strongly influenced by land use, topography and other local features not resolved by the horizontal resolution used in GCMs ([Gutmann et al. 2011], [Boé et al. 2009]). On the other hand, atmospheric circulation is assumed to be better simulated by GCMs, since it is characterised by large-scale variations ([Plaut and Simonnet 2001]). It is also less dependent on surface influences, in particular when considering upper levels, for example the geopotential height at 500 hPa.

Furthermore, the atmospheric circulation simulated by GCMs is used in many climatological studies and as a forcing for many downscaling methods. For example, GCMs are necessary inputs as boundary conditions for Regional Climate Model simulations ([Zorita and von Storch 1999]). They are also used as a predictor variable for statistical downscaling methods ([Anagnostopoulou et al. 2008], [Brinkmann 2000], [Enke and Spekat 1997]). But, whereas statistical and dynamical downscaling methods attempt to give more precise results at the surface than GCMs, they are not able to correct the biases in the atmospheric circulation simulated by GCMs ([Fettweis et al. 2011a], [Yoshimori and Abe-Ouchi 2011]). Thus, the reliability and the correctness of the GCM-based general circulation are very important given that they are essential assumptions for the use of this circulation in downscaling methods ([Wilby and Wigley 2000], [Yarnal et al. 2001]). Therefore, it is essential to analyse and evaluate the general circulation simulated by GCMs.

Circulation type classifications are efficient tools to evaluate GCM-based circulations ([Pastor and Casado 2012],[Anagnostopoulou et al. 2009], [Schuenemann and Cassano 2009], [Zorita et al. 1995], [Kysely and Huth 2006], [Bardossy and Caspary 1990], [Demuzere et al. 2009], [Huth 2000]) and to analyse in detail projected changes in the future circulation ([Schuenemann and Cassano 2010]). Indeed, these classifications allow a more precise analysis of the general circulation by considering each circulation type separately ([Bardossy et al. 2002]). Therefore, circulation type classifications have the advantage over simple statistics, which are often based only on the average and the standard deviation of the present day conditions.

Since GCMs do not reproduce the daily observed climate but try to simulate as well as possible the mean climatic state and its variability over a long period, it is not possible to analyse the outputs of the models day by day. It is for this reason that monthly or seasonal means over many years are usually used to compare GCM outputs with reference datasets such as reanalyses ([Franco et al. 2011], [Walsh et al. 2008]). But, these approaches ignore the variability of the atmospheric circulation and of the associated weather conditions at the surface, which can be observed on daily to weekly time scales ([Casado and Pastor 2012]). Circulation type classifications avoid this problem by grouping and averaging similar daily circulation situations together through minimising the within-type variability. This therefore allows a precise and subtle analysis of circulation patterns, since each relatively homogeneous type can be examined separately. Moreover, given that the principle of any classification is to characterise the diversity of a dataset, circulation type classifications better focus on the ability of GCMs to reproduce the variability of the atmospheric circulation over a region. This is considered by [Overland et al. 2011] as the first step in the procedure for selecting a subgroup consisting of the best GCMs. A reliable simulation of the variability of the circulation is essential, since changes in this variability, meaning changes in circulation patterns, affect the surface climate conditions ([Casado and Pastor 2012], [Stoner et al. 2009]). In particular, extreme weather conditions and their impacts are usually observed under extreme circulation situations, enhancing the need for simulations able to reproduce the diversity of the circulation. Finally, circulation type classifications have a high computational efficiency, which allows the evaluation of a large number of GCMs ([Boé et al. 2009]).

Taking all this into account, we used the circulation type classification developed by [Fettweis et al. 2011b] over Greenland to compare the daily geopotential height at 500 hPa simulated by GCMs with three reanalyses for the current climate (1961-1990). With the aim of studying the GrIS surface mass balance, we mainly focused our comparison on the summer months (JJA, for June, July and August). We chose these months because the atmospheric circulation has a great impact, in addition to precipitation ([Schuenemann and Cassano 2009]), on the surface melt, which occurs essentially during summer. Indeed, the surface melt is strongly influenced by the temperature, which is highly correlated to the geopotential height, according to [Fettweis et al. 2011b]. An evaluation of the GCM-based general circulation during the winter (DJF, for December, January and February) is, however, provided in the Supplementary Material. The comparison between the datasets is based on differences in the frequency distribution of each circulation type between the GCMs and the reanalyses and on an analysis of the intraclass variability. Moreover, this approach is extended to future climate simulations to study the projected changes in the atmospheric circulation under warmer climates over Greenland.

2 Data

As proposed by [Fettweis et al. 2011b], we used the geopotential height at 500 hPa as the input variable of the circulation type classification for evaluating the general circulation over Greenland. GCMs from the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel dataset and its successor CMIP5 prepared respectively for the IPCC assessment reports AR4 ([Randall et al. 2007]) and AR5 were used in this study to examine whether there has been an improvement between the CMIP3 GCMs and their new CMIP5 version. All the GCMs for which we could obtain geopotential height data were used here. Since the geopotential height was not a requested variable in CMIP3 and is only a second priority variable in CMIP5, daily data for only a few GCMs could be retrieved. For the CMIP3 GCMs, all monthly data and the outputs of BCCR come from the CMIP3 database (see Table 1). The other output data were downloaded directly from the modelling centre databases. For the CMIP5 GCMs, all outputs were downloaded from the CMIP5 platform.

In order to evaluate the ability of GCMs to simulate the 20th century climate, daily and monthly mean summer (June, July and August) 500 hPa geopotential heights (referred to hereafter as Z500) were downloaded for the period 1961-1990. The monthly data were used as a basis for interpreting the results of the classification. The scenarios representing the current climate conditions are called 20C3M (20th Century Climate in Coupled Models) for CMIP3 and Historical for CMIP5. For the CMIP5 future projections, we used two Representative Concentration Pathway (RCP) experiments: the mid-range experiment RCP4.5 projecting a radiative forcing of 4.5 W/m^2 till 2100 and the pessimistic experiment RCP8.5 simulating a radiative forcing of more than 8.5 W/m^2 till 2100 ([Moss et al. 2010]). Calculations were made for the first run (run1 for CMIP3 and r1i1p1 for CMIP5) of each GCM.

The GCM outputs were compared to three reanalysis datasets: the NCEP/NCAR Reanalysis from the National Centers for Environmental Prediction - National Center for Atmospheric Research ([Kalnay et al. 1996]), the ERA-40 Reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF)([Uppala et al. 2005]) and the Twentieth Century Reanalysis version 2 (20CR)([Compo et al. 2011]) from the NOAA ESRL/PSD (National Oceanic and Atmospheric Administration Earth System Research Laboratory/Physical Sciences Division). More recent reanalysis datasets such as ERA-Interim, NCEP-DOE or MERRA are not used here, since they start around 1979 while the GCM simulations for the current climate go till 2000 (CMIP3) and 2005 (CMIP5). The overlapping period for the current climate evaluation would not be long enough (i.e. at least 30 years) to give robust results that are less influenced by the natural variability of the circulation.

As the reanalyses and GCM outputs have different spatial resolutions (Table 1), the daily data used for the classification were linearly interpolated on a regular grid of 100 km resolution. As proposed by [Fettweis et al. 2011b], an area of 1400 km by 2700 km covering Greenland (centred on 72^{o}N 40^{o}W) was selected as the classification domain (see Figure 1). They showed that this domain is the most appropriate to study the atmospheric circulation over the GrIS with the methodology used here.

3 Methodology

Many classification methods have been developed during the last few decades for climatic or meteorological purposes ([Huth et al. 2007]). Their aim is to group meteorological situations on the basis of atmospheric circulation (circulation type classifications) or according to surface weather elements (weather type classifications) into some distinct patterns in order to characterise the climatic conditions of the studied region ([El-Kadi and Smithson 1992], [Yarnal et al. 2001], [Huth 2000], [Philipp et al. 2010]). The first classifications were manual and an operator allocated each situation to the most similar type. Most of these methods have now been automated, but they remain partially subjective, since the types are predefined; these methods are therefore considered as hybrid. Many automatic methods, where the types are defined through an algorithm and not by the user, are also available ([Philipp et al. 2010]). They often use a principal component analysis ([Casado et al. 2009], [Huth 2000]), the correlation ([Lund 1963]), the root mean square deviation ([Kirchhofer 1973]) or the Euclidean distance ([Philipp et al. 2007]) between the circulation situations to quantify their similarities. Then, a clustering technique such as K-means, Ward's method, average linkage, the centroid method or a leader algorithm is used to find the types and assign each situation to one of these types ([El-Kadi and Smithson 1992], [Kalkstein et al. 1987], [Huth et al. 2007]). In the last few years, more complex methods such as self-organising maps have been developed ([Schuenemann and Cassano 2009]). Nevertheless, a comparison of many of these methods shows that no particular method can be considered as being better than the others ([Philipp et al. 2010]).

Here, we used two indices to characterise the similarity between the pairs of daily circulation situations (i.e. daily mean geopotential height at 500 hPa), according to which the circulation situations were assigned to particular circulation-type classes. The first index, impacted by the geopotential height of the circulation situations, is based on the normalized Euclidean distance (referred to hereafter as DIST) between the two Z500 surfaces for each pair of days, as defined by [Fettweis et al. 2011b]. So, two situations with a similar geopotential height but slightly different patterns can be grouped together in contrast to two situations presenting the same pattern but at different mean geopotential heights. The aim of this paper was to evaluate the GCM circulation as a forcing for Regional Climate Models (RCMs) over Greenland. Thus, we needed to take into account the geopotential height, since this is highly correlated to the atmospheric temperature, which affects the melting rate simulated by the RCMs ([Fettweis et al. 2011b]). However, the influence of the mean geopotential height introduces artefacts in some specific cases, as we will see. To overcome this drawback, a second index is used. This index, evaluating only the pattern (i.e. the position of high and low pressures, regardless of the gradient strength) of the Z500 surface, is defined as the Spearman rank correlation coefficient (referred to hereafter as RANK) for all pairs of situations. As argued by [Vautard and Yiou 2009], who used this coefficient to find analogues, the advantage of using the Spearman rank correlation rather than the linear correlation coefficient is that it avoids the influence of outliers on the index. This means that the Spearman rank correlation coefficient between two situations with similar patterns but with different gradient strengths is higher than their linear correlation coefficient. However, two parallel but distant Z500 surfaces are considered as similar with the correlation-based index because they have the same pattern. But, if the Z500 surfaces are parallel, this means that the temperature of the troposphere below 500 hPa is different and so, these two Z500 surfaces will not have the same impact on the surface climate or as forcing fields for an RCM. Moreover, the gradient strength difference between two surfaces with a similar pattern is not taken into account ([Philipp et al. 2007]). For example, a strong and a weak anticyclone will be grouped together using RANK, regardless of the strength of the anticyclones, whereas they are treated as separated types by DIST. However, this approach offers the advantage of being independent of a warming of the atmosphere.

Once the index is calculated, the circulation types are determined through an automatic circulation type classification developed with the aim of linking the atmospheric circulation over Greenland to the GrIS surface melt and described by [Fettweis et al. 2011b]. This classification is considered as a leader algorithm method ([Philipp et al. 2010]). That means that the first class is defined by the situation (called hereafter the reference situation) that counts the most similar situations, two situations being considered as similar if their index is above a given threshold. The second class is built in the same way on the basis of the remaining situations, and so on for all classes. Since the number of classes is fixed by the user, the threshold above which two situations are considered as similar is decreased class by class, given that the similarity indices reach 1 for two identical situations and decrease with increasing dissimilarity. This avoids very dissimilar sizes between the first and the last classes. When the requested number of classes is built and the number of unclassified situations is below a threshold (fixed here at 1%), these remaining situations are assigned to the last class. This means that this class can be dominated by one circulation pattern, but that it can take into account some very dissimilar patterns. In order to optimise the percentage of explained variance and so to reduce the within-type variability ([Philipp et al. 2010]), the classification scheme is repeated many times with various decrement and threshold values.

Since the classification of circulation types used here is an automatic one, the circulation types are derived from the classification process and not predefined by the user. This implies that different datasets give different classification results for the same period. So, any comparison between the circulation types of these datasets is impossible. To avoid this problem, [Huth 2000] suggests "projecting" the types of one dataset, considered as the reference, onto the other datasets. Here, we used the ERA-40 reanalysis as the reference dataset, but as shown by [Brands et al. 2012], NCEP/NCAR and ERA-40 present very similar circulation patterns, so that it makes almost no difference whether one or the other dataset is used as the reference. Moreover, most other studies evaluating the GCM-based circulation over Greenland or the Arctic region have used ERA-40 as the reference dataset ([Walsh et al. 2008], [Schuenemann and Cassano 2009]). As a benchmark for GCMs, NCEP/NCAR and 20CR are compared here to ERA-40 in the same way as for GCMs.

The projection of the reference types onto a dataset consists of classifying the situations of this dataset using the same parameters that define the classes derived from the reference dataset. In our case, each class is defined by its reference situation and its index threshold. These two parameters are imposed on the GCM and the other reanalysis (NCEP/NCAR and 20CR) datasets to assign the situations to the classes, so that the types remain exactly the same for all GCMs, experiments and periods. This allows an easy comparison type by type, solely on the basis of differences in the frequency of the classes between the datasets. Since the unclassified situations are assigned to the last class, the more its frequency is overestimated by a GCM, the more this GCM fails to reproduce the observed types. For future climate projections, this also means that if new circulation types appear due to climate change, these situations will fall into this class.

We used the RMSE (root mean square error) between the ERA-40 and the GCM frequencies as the synthetic index for comparison. However, although the parameters defining the classes are identical for all datasets, the distribution of the situations within the classes can differ from one dataset to another. This means that biases or circulation changes due to global warming can affect the distribution of the situations within the GCM classes, particularly for RANK, since its classes do not depend on the geopotential height. To highlight intraclass distribution differences between the ERA-40 and the GCM classes, a two-sample

Kolmogorov-Smirnov statistic (referred to hereafter as the KS-test) was calculated for each class. Finally, to ensure that our results were not influenced by the projection, an automatic classification was also carried out for some GCMs and the obtained types were projected onto the ERA-40 dataset, as proposed by [Huth 2000].

Using DIST to classify the daily Z500, [Fettweis et al. 2011b] showed that eight classes are sufficient to represent the main circulation types observed over Greenland during summer and that a domain limited approximately to the Greenland coasts gives the best results for NCEP/NCAR. The circulation types obtained for the reference classification using ERA-40 daily mean Z500 data for June, July and August for the period 1961-1990 can be divided into three categories: anticyclonic, cyclonic and zonal flow types (see Fig. 1). The anticyclonic (corresponding to a negative North Atlantic Oscillation (NAO) index) and the cyclonic (corresponding to a positive NAO index) categories are both divided into two types. The first type shows a weak gradient, and thus a weak ridge (Class 3) or trough (Class 2), and is relatively frequent (around 20%). The second type has a stronger gradient and is therefore less frequent (Class 7 showing a well marked anticyclone over southern Greenland and Class 5 presenting a broad trough). Anticyclonic (resp. cyclonic) types favour on average warmer (resp. colder) atmospheric conditions compared to the seasonal mean, as shown by [Fettweis et al. 2011b]. Class 1 groups the intermediate circulation situations showing no clear anticyclonic or cyclonic curvature and is therefore close to the mean pattern over the period. In the zonal flux category, Class 4 is characterised by a strong north-west to south-east gradient (westerly flow), whereas the other zonal type (Class 6) shows a reversed situation with a higher Z500 in the north than in the south of Greenland, inducing an easterly flow. The last type (Class 8, accounting for 0.7% of the sample) is composed of both a circulation type showing a strong westerly flow and the unclassified situations, which are very heterogeneous. As shown in Fig. 2, RANK gives patterns very different from DIST. The RANK types highlight flow patterns (with both positive and negative anomalies for each class) rather than cyclonic and anticyclonic patterns, as typed by DIST. As we will see later, the interpretation of the frequency biases of the GCMs for these types is much more difficult than for DIST.

4 Evaluation of 20th century circulation types

4.1 JJA mean Z500

Before comparing the frequency differences for each circulation pattern between the GCMs and the reanalyses, it is important to evaluate the ability of the GCMs to reproduce the JJA mean Z500 (referred to hereafter as $Z500_{JJA}$) over Greenland and its pattern for the current climate (1961-1990). Indeed, since DIST is influenced by the geopotential height, a GCM showing a strong $Z500_{JJA}$ anomaly also gives classification results very different from those of the reanalyses. Moreover, anomalies in the mean geopotential height suggest that the simulated atmosphere could be too warm or too cold, bearing in mind that temperature and geopotential height are positively correlated. So, a GCM presenting a high $Z500_{JJA}$ anomaly cannot be reliably used as a forcing input for downscaling methods. Finally, if a GCM is not able to simulate correctly the current climate, its ability to simulate future projections might be questionable. Some studies ([Masson and Knutti 2011], [Reifen and Toumi 2009]) have shown that the consistent results of one GCM over a given period cannot be considered as a guarantee of good results for other periods. However, it is likely that good matching GCMs over the 20th century will give more realistic future projections than GCMs that fail to reproduce the current circulation ([Yoshimori and Abe-Ouchi 2011], [Casado and Pastor 2012]).

Figure 3 shows the $Z500_{JJA}$ anomaly with respect to ERA-40 over Greenland for the reanalyses and the GCMs over the 1961-1990 period. The root mean square error between each GCM and ERA-40 is listed to quantify the differences in $Z500_{JJA}$. We can immediately see a very close similarity between the three reanalyses, despite the fact that 20CR slightly overestimates $Z500_{JJA}$. It should be remembered that only the surface pressure is assimilated in the 20CR reanalysis in contrast to ERA-40 and NCEP/NCAR, which also use upper air data. So, we can expect that 20CR will give worse results, especially in the upper atmosphere. The differences between the GCMs and ERA-40 are generally much larger. It appears that the $Z500_{JJA}$ anomaly is very different from one GCM to another and that it can be negative as well as positive, so that no general tendency can be observed, as already shown by [Walsh et al. 2008] for CMIP3 models over the Arctic region. Nevertheless, the comparison cannot be made only on the basis of the RMSE and the mean differences, as they do not take into account the ability of the GCMs to reproduce the mean pattern. As described by [Franco et al. 2011], this pattern is characterised by a south-west to north-east flow over the Baffin Bay turning to an eastward circulation over the GrIS except for southern Greenland, where the circulation remains from the south-west. When looking further into this $Z500_{JJA}$ pattern, only a relatively few (about one fourth) of the GCMs can be considered as being able to reproduce this pattern (for example, HadGEM1, IPSL4, HadGEM2 and MIROC5). The other GCMs show too weak of a north-south gradient (for example, BCCR or CNRM), an excessive ridge over Greenland (for example, IPSL5-LR and MRI) or have no realistic pattern (for instance, GISS-E2-R). Some GCMs such as BCC, CanESM2, MPI-LR or NorESM1 present artefacts in the isohypses over Greenland (probably due to the ice sheet topography) but, in general, their patterns are similar to those of the reanalyses.

When comparing the CMIP3 and the CMIP5 versions of GCMs, we can observe that only in the case of CCCma47 and CCCma63, the $Z500_{JJA}$ anomalies are larger than in the CMIP5 version (CanESM2). For HadGEM and HadCM3, the anomalies are similar and IPSL4 shows a pattern closer to that of the reanalyses and a lower $Z500_{JJA}$ anomaly than its new versions (IPSL5-LR, -MR and IPSL-CM5B-LR).

For the detailed analysis on a daily time-scale of the GCM-based circulation with the help of the circulation type classification, we used all GCMs (CMIP3 and CMIP5) for which daily Z500 outputs are available.

4.2 Classification results

The class by class frequency distribution for DIST shows that NCEP/NCAR generally gives frequencies closer to the ERA-40 frequencies than most of the GCMs (see Figure 4). The good agreement between NCEP/NCAR and ERA-40 is confirmed by [Casado et al. 2009], who compared the results of a classification of both reanalyses for winter in Europe. For 20CR, the differences with regard to ERA-40 are larger than and of the same order (in absolute value) as those between ERA-40 and the best matching GCMs. For this reanalysis as well as for the GCMs, the frequency biases reflect the $Z500_{JJA}$ anomalies discussed above. Indeed, classes 3 and 7 (anticyclonic classes with a positive anomaly, see Fig. 1) are overrepresented by the GCMs presenting a positive $Z500_{JJA}$ anomaly, which is the case for most of them. This is particularly marked for the GCMs showing an anticyclonic ridge over Greenland (for example, IPSL5-LR and MIROC-E). On the other hand, the GCMs presenting a negative $Z500_{JJA}$ anomaly underestimate the frequency of these classes. Of course, for the cyclonic classes (2 and 5), the comparison is analogous. Since Class 1 has a small $Z500_{JJA}$

anomaly, no clear trend can be highlighted for the GCMs. The westerly flow type (Class 4) and the easterly flow type (Class 6) are underrepresented by (nearly) all GCMs. So, it seems that these types are more difficult to simulate than the more basic anticyclonic and cyclonic types. Finally, most of the GCMs overestimate the frequency and the variability of the last class, which includes the non-classified days. This shows that most of the GCMs simulate too many days with patterns that are very different from the 7 reference ERA-40 based patterns, but also that this class is not dominated by new circulation types (which would induce a lower standard deviation in this case).

RANK confirms the results obtained on the basis of DIST (see Table 2 and Supplementary Material ESM-Fig. 1). Indeed, the GCMs showing the closest frequency distribution to ERA-40 are the same for both classifications. Moreover, some general trends can be highlighted. Classes 4 and 6 are underrepresented in most GCM datasets, while classes 5 and 8 are overrepresented. The other classes show no clear tendency. Some GCMs largely over- or underestimate some classes, simulating half or twice the expected frequency. In contrary to DIST, it is difficult to link the frequency biases of the GCMs to their $Z500_{JJA}$ biases. When considering the KS-test, it appears that only CanESM2 shows similar intraclass distributions for most of its classes with regard to the corresponding ERA-40 intraclass distributions. This GCM also has the lowest $Z500_{JJA}$ bias. The other GCMs have significantly different intraclass distributions for (nearly) all classes. This means that the $Z500_{JJA}$ bias is not only due to the over- or underestimation of the frequency of some circulation types, but that it affects the whole circulation. This is also confirmed by the lower RMSE values and higher number of classes with a significantly different intraclass distribution for RANK than for DIST. Indeed, the higher RMSE and the lower number of classes with a significantly different intraclass distribution for DIST can be explained by the influence of the geopotential height. The differences between the two classifications also highlight differences between the GCMs. For example, the IPSL5 GCMs show high RMSE values for both DIST and RANK, indicating that their frequency biases highlighted with DIST are indeed due to biases in the frequency distribution of the circulation patterns (i.e. an overrepresentation of the anticyclonic types). By contrast, MIROC-E and MIROC-EC present a very high RMSE for DIST and a much lower RMSE for RANK. This means that the frequency biases of these GCMs for DIST are rather due to their $Z500_{JJA}$ bias than to an important over- or underestimation of some circulation patterns. But let us remember that a $Z500_{JJA}$ bias is likely to induce temperature biases in the hosted RCM (according to [Fettweis et al. 2011b]), while a frequency bias will impact the occurrence of the number of warm and cold events during summer.

4.3 Persistence of the circulation types

The persistence of a circulation type is calculated as the mean number of consecutive days grouped in this type. In general, it appears that the persistence is overestimated and that the persistence biases are related to the frequency biases (Fig. 5). Indeed, the two classes which show too low a persistence for most GCMs are classes 4 and 6, which are also underrepresented by most GCMs. Moreover, the GCMs overestimating the anticyclonic type frequencies (for example, IPSL5-LR or MIROC-EC) also simulate a higher persistence for these types (generally about one to two days). This is logical, since if a type is more frequent, it is more likely to have a higher persistence. An analogous explanation can be held for the GCMs overrepresenting the cyclonic types. On the other hand, the persistence of the types that are underrepresented is generally close to that of ERA-40, while one could expect that this persistence would also be underestimated. This anomaly might be due to the general overestimation of persistence by the GCMs. As shown in the Supplementary Material (ESM-Fig. 2), this overestimation of persistence also appears for RANK, where the biases are lower, as in the case of frequency biases.

It is important to note that the observations made here are similar when using NCEP/NCAR as the reference dataset instead of ERA-40 (see Supplementary Material ESM-Fig. 3, 4 and ESM-Table 1). Moreover, as indicated by [Huth 2000], the projection of the types of one dataset onto the other should be done in both directions to ensure that the results are not influenced by the projection itself. This was done in the present study for 5 GCMs (CanESM2, IPSL5-LR, MPI-LR, MRI and NorESM1). The automatic classification was run on the Historical (1961-1990) dataset of these GCMs and the resulting types imposed onto both ERA-40 and NCEP/NCAR. The RMSE over the frequency differences and the number of classes with a significantly different intraclass distribution (based on the KS-test) were found to be of the same order as presented above (see Supplementary Material ESM-Table 2 and Table 2). Moreover, when highlighting the IPSL5-LR classification using DIST, it counts 5 anticyclonic types, most of which are underrepresented by ERA-40 and NCEP/NCAR (see

Supplementary Material ESM-Fig. 5 and ESM-Table 3). This confirms the observations made before that IPSL5-LR over-simulates anticyclonic situations.

Finally, we compared the classification results using ERA-40 as the reference dataset for the 5 first runs (from r1i1p1 to r5i1p1) for CanESM2. It appears that the spread is quite low (with an RMSE varying between 2.15 and 4.12 for DIST and between 2.44 and 3.91 for RANK). So, this suggests that the differences between the runs of the same GCM are lower than those between the GCMs. This might be due to systematic errors or to the parametrisation, which remains almost the same for a particular GCM. This is confirmed by the $Z500_{JJA}$ patterns and biases, which are often similar for GCMs from the same institute, when compared to GCMs from different research centres. In this way, the increased resolution for some GCMs (CCCma47 - CCCma63, IPSL5-LR - IPSL5-MR and MPI-LR - MPI-MR) does not seem to improve nor to deteriorate significantly the ability of these GCMs to reproduce the observed atmospheric circulation. When comparing the performance of CMIP3 and CMIP5 GCMs, no improvement was detected.

5 Future projections of the circulation

In this section, we will focus on some of the CMIP5 GCMs that best simulate the current climate, on the basis of both the DIST and RANK frequency RMSE and KS-test values: BCC, CanESM2, MPI-MR and NorESM1. This selection of the best matching GCMs is in agreement with the conclusions of [Overland et al. 2011] and [Walsh et al. 2008], who observed that, in relation to the Arctic, the most reliable GCMs are those most sensitive to climate change. Moreover, the general conclusions of this section are also valid for the other GCMs used previously. The future experiments selected here are the Representative Concentration Pathways RCP4.5 and RCP8.5 described in section 2, which can be considered as the mid-range and the upper limit experiment, respectively. In order to perform the classification and to apply the same approach as for the current climate, the future projections are split into three 30-year periods: 2011-2040, 2041-2070 and 2071-2100.

First, let us analyse the results obtained with RANK (see Table 3 and Figure 6). It appears that there are no significant or systematic circulation changes through the three future periods or for the two experiments. It is true that there are some small changes through the three periods for some classes. However, on the one hand, these changes account for only 2 to 5% between the first and the last future period and on the other hand, they are lower than or are of the same order as the frequency biases between the GCMs and ERA-40 for the current climate. This means that the GCMs simulate neither new circulation patterns nor significant frequency changes under climate change conditions. Persistence also does not show any significant changes through the future periods with regard to the Historical experiment (see Table 4). Despite the interdependence between frequency and persistence, it is possible to observe persistence changes without frequency changes, but this is not the case here. However, the KS-test values show a strong increase for all classes, showing that the intraclass distribution calculated on the basis of the daily mean Z500 becomes increasingly different under climate change conditions. A more detailed analysis shows that the mean Z500 of all classes increases towards 2100. This is confirmed by the simulated $Z500_{JJA}$ (Fig. 7), which shows a progressive increase induced by the warming of the atmosphere through the three future periods compared to the current climate (Section 4.1). This $Z500_{JJA}$ increase is consistent with a warming over the whole North America - North Atlantic - Europe domain (not shown). Of course, the increase is more pronounced for RCP8.5 than for RCP4.5. However, it is interesting to observe that the $Z500_{JJA}$ pattern remains the same for the two future experiments compared to the current climate for all three periods, confirming that there is no significant change in the circulation type frequencies. This observation is in contradiction with the results obtained by [Franco et al. 2011]. They showed for CMIP3 GCMs a stronger mean Z500 increase over the northern part of Greenland. This probably means that the warming and the associated Z500 increase is spatially more homogeneous for CMIP5 GCMs than for CMIP3 GCMs. Note that [Franco et al. 2011] worked over the whole year and that the CMIP3 future experiments (A1B, A2 and B1) are difficult to compare with the CMIP5 future experiments since they are defined differently.

On the other hand, the DIST results show significant and systematic frequency changes in the circulation types. The most important changes are a rarefaction of the cyclonic types and a strong increase in the anticyclonic type frequencies. Nevertheless, these frequency changes are an artefact associated to the warming of the atmosphere due to the influence of the geopotential height itself on DIST. For the future climate, the Z500 increase is strong enough so that the difference in geopotential height between the future Z500 surfaces and the ERA-40 surfaces becomes dominant, to the detriment of the pattern. In this case, the classes can no longer be interpreted as circulation types. To avoid this artefact, we removed from each future daily Z500, the $Z500_{JJA}$ increase between the Historical (1961-1990) and the considered future period, before applying DIST again. The aim of this reasoning was to verify whether DIST gives the same results, i.e. no systematic circulation changes, as RANK, when it is not influenced by the Z500 increase. Removing the $Z500_{JJA}$ increase is justified, since the KS-test for RANK and the future $Z500_{JJA}$ pattern give some evidence that the Z500 increase is similar for all classes (see also Supplementary Material ESM-Table 4 for the differences in the class means between the future experiments and the Historical experiment). The results obtained for DIST after removing the $Z500_{JJA}$ increase confirm that the GCMs do not simulate significant changes in the circulation type frequencies (see Table 3 and Supplementary Material ESM-Fig. 6).

6 Results for winter

When applying the method explained here to the winter months (December, January and February), it appears that the general conclusions are the same as for the summer. The ranking of the best matching GCMs is only slightly different, since some good matching GCMs for summer give worse results for winter (see Supplementary Material ESM-Fig. 7, 8 and ESM-Table 5). For example, BCC and CanESM2 strongly overestimate the frequency of the cyclonic classes, while MIROC5 overrepresents the anticyclonic types. It is also interesting to note that some GCMs that match worse for the summer give better results for the winter (IPSL5-LR and -MR, MIROC-E and -EC). The other GCMs fail to reproduce the winter ERA-40-based circulation. As for the summer, both RANK and DIST (after removing the $Z500_{DJF}$ increase) show that none of the GCMs simulates significant circulation changes for the future compared to the current climate.

7 Conclusion

We evaluated the Z500 circulation simulated by the CMIP3 and CMIP5 GCMs over Greenland with the help of a circulation type classification. Two different indices were used: the Euclidean distance and the Spearman rank correlation coefficient. These two indices give very different circulation types, since the first is influenced by the differences in geopotential height between the daily situations, while the second takes only the circulation pattern into account. It is interesting to observe that the best matching GCMs for the current climate (for summer: HadGEM1, IPSL4, BCC, BNU, CanESM2, MIROC5, MPI-MR and NorESM1 and for winter: HadGEM1, IPSL4, BNU, HadGEM2, MPI-LR, MPI-MR and NorESM1) are the same for both indices. This shows the independence of the results in respect to the index used.

For the current climate, some major differences in the frequency of the circulation types between the GCM-based circulation and ERA-40 were highlighted for most GCMs. Obviously, these GCMs have difficulty in reproducing the observed circulation over Greenland during summer and winter. Indeed, despite the ability of most GCMs to reproduce the observed circulation types, the differences between them and ERA-40 are much higher than those between NCEP/NCAR and ERA-40. This discrepancy gives an idea of the uncertainties of the GCM-based geopotential height data over Greenland. Through the strong relationship between the atmospheric circulation and other variables such as temperature, precipitation and wind, the frequency biases of the circulation types have important implications for the reliability of these variables, as shown by [Schuenemann and Cassano 2009] for precipitation. The frequency and persistence biases show the difficulty for the GCMs in reproducing the variability of the atmospheric circulation. In particular, the study of rare and extreme circulation types might be risky, since these conditions will probably not be well simulated. Our results for the current climate join the conclusions of [Stoner et al. 2009] and [Casado and Pastor 2012], who showed that some GCMs give more realistic results than others, but that there is no one GCM that is systematically and significantly better than the others. As stated by [Overland et al. 2011], the selection of a particular GCM depends on the application, but some GCMs might be more useful than others over Greenland, particularly as a forcing for Regional Climate Models, which need GCM-based circulation forcing at high temporal resolution.

We also showed that the relationship between the frequency biases and the $Z500_{\mathrm{JJA}}$ bias is different from one GCM to another. For some GCMs, the $Z500_{\mathrm{JJA}}$ bias seems to affect all circulation types in more or less the same way because the $Z500_{\mathrm{JJA}}$ bias is induced by an atmospheric temperature bias. For other GCMs, the frequency biases of some classes (e.g. anticyclonic types) are so important, that they induce a $Z500_{\mathrm{JJA}}$ bias. On the one hand, this means that it is very dangerous to simply remove the mean bias of a GCM variable before using it. On the other hand, it confirms the need for a GCM evaluation on a daily to sub-daily timescale before using the GCMs at this timescale, for example as a Regional Climate Model forcing. Circulation type classifications are an efficient tool to achieve such an evaluation, since they allow us to consider the ability of the GCMs to reproduce the diversity of the circulation types as well as the variability of the atmospheric circulation on a daily timescale, which is not possible with monthly or seasonal mean approaches. For example, the underestimation of some types (classes 4 and 6) is not detected using $Z500_{JJA}$. Moreover, it is impossible to know from $Z500_{JJA}$ whether a GCM showing an anticyclonic anomaly overrepresents strongly Class 3 or slightly Class 7, despite major differences in the impact of these classes on variables such as temperature or precipitation. In general, $Z500_{JJA}$ gives no quantitative information on the over- or underestimation of the different types and consequently their persistence, in spite of the influence of the persistence on blocking conditions, for example.

For the future projections, RANK suggests almost no circulation changes. This is confirmed by DIST after removing the $Z500_{JJA}$ increase between the future period and the current climate. In this case, the removal of the $Z500_{JJA}$ increase is justified, since it affects all circulation types in a similar way, as evidenced by RANK. The absence of circulation changes means that the changes in other variables such as temperature and precipitation are due to changes in the intraclass variability of these variables. This has been pointed out by [Schuenemann and Cassano 2010], who showed that the most important projected precipitation changes are due to changes in the intraclass variability of precipitation, and the KS-test for RANK shows a strong Z500 increase for all classes, which is explained by the warming of the region. This also means that we could gain a good idea of the future climate changes simply by using the ERA-40 (or NCEP/NCAR) circulation and only changing the temperature and its associated variables (such as humidity, precipitation, cloudiness, etc.), but not the regional wind, since this depends on the circulation patterns and therefore should not change significantly due to global warming. Nevertheless, we cannot conclude that the projected warming over the region does not imply changes in the frequency distribution of the circulation types; nor can we conclude that the GCMs are not able to simulate frequency changes. However, according to [Fettweis et al. 2011b] and [Hanna et al. 2009], it should be noted that the recent JJA warming in the 2000s over Greenland seems to result from changes in circulation patterns with more anticyclonic conditions than over the last few decades, favouring southerly warm air advection over the western part of Greenland. These more anticyclonic conditions are related to a strong decrease in the NAO index, as it appears on Fig. 8. On the one hand, GCMs have obvious difficulty in simulating similar conditions. On the other hand, the projected absence of NAO changes towards 2100 does not allow us to conclude whether the NAO changes observed over the last few years should be attributed to climate variability, or if these changes are due to global warming.

Another important result is that the different runs of the same GCM gave similar results for atmospheric circulation. This means, on the one hand, that it does not change significantly the results if another run is used instead of r1i1p1, and on the other hand, that different runs of one GCM can only be used to quantify the uncertainties related to the parametrisation of that particular GCM. One cannot use different runs of only one GCM to gain an idea of the spread of the values for a given experiment. Moreover, the spread of the GCM simulations for one experiment gives an idea of the uncertainty over this experiment and, as already advised by [Overland et al. 2011], it is necessary to work with several GCMs to gain an idea of the extent of this uncertainty.

Acknowledgements We acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP Working Group on Coupled Modelling (WGCM) for their role in making available the WCRP CMIP3 multi-model dataset. Support for this dataset is provided by the Office of Science, U.S. Department of Energy.

For their role in producing, coordinating, and making available the CMIP5 model output, we acknowledge the climate modelling groups (see Table 1), the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM), and the Global Organization for Earth System Science Portals (GO-ESSP).

We also wish to thank the research centres that publish their data, especially data not contained in the CMIP databases.

NCEP/NCAR Reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/.

The ECMWF ERA-40 Reanalysis data used in this study were obtained from the ECMWF Data Server (http://www.ecmwf.int).

Support for the Twentieth Century Reanalysis (20CR) Project dataset was provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, by the Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office.

Finally, we wish to thank the ISLV Editing and Translation Services for their support in improving the language of this manuscript.

References

[Anagnostopoulou et al. 2008] Anagnostopoulou C., Tolika K., Maheras P., Kutiel H., Flocas H. (2008) Performance of the general circulation HadAM3P model in simulating circulation types over the Mediterranean region, International Journal of Climatology, 28, 185-203

[Anagnostopoulou et al. 2009] Anagnostopoulou C., Tolika K., Maheras P. (2009) Classification of circulation types: a new flexible automated approach applicable to NCEP and GCM datasets, Theoretical and Applied Climatology, 96(1-2), 3-15

[Bardossy and Caspary 1990] Bardossy A., Caspary H.-J. (1990) Detection of climate change in Europe by analyzing European atmospheric circulation patterns from 1881 to 1989, Theoretical and Applied Climatology, 42, 155-167

[Bardossy et al. 2002] Bardossy A., Stehlik J., Caspary H.-J. (2002) Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules, Climate Research, 23, 11-22

[Boé et al. 2009] Boé J., Terray L., Martin E., Habets F. (2009) Projected changes in components of the hydrological cycle in French river basins during the 21^{st} century, Water Resources Research, 45, doi: 10.1029/2008WR007437

[Box et al. 2010] Box J., Cappelen J., Decker D., Fettweis X., Mote T., Tedesco M., van de Wal R. (2010) Greenland. In: Arctic Report Card 2010, http://www.arctic.noaa.gov/reportcard

[Box et al. 2012] Box J., Fettweis X., Stroeve J., Tedesco M., Hall D., Steffen K. (2012) Greenland ice sheet albedo feedback: thermodynamics and atmospheric drivers, The Cryosphere Discussions, 6, 593-634, doi:10.5194/tcd-6-593-2012

[Brands et al. 2012] Brands S., Gutierrez J., Herrera S., Cofiño A. (2012) On the use of reanalysis data for downscaling, Journal of Climate, 25, 2517-2526

[Brinkmann 2000] Brinkmann W. (2000) Modification of a correlation-based circulation pattern classification to reduce within-type variability of temperature and precipitation, International Journal of Climatology, 20, 839-852

[Casado et al. 2009] Casado M., Pastor M., Doblas-Reyes F. (2009) Euro-Atlantic circulation types and modes of variability in winter, Theoretical and Applied Climatology, 96, 17-29

[Casado and Pastor 2012] Casado M., Pastor M. (2012) Use of variability modes to evaluate AR4 climate models over the Euro-Atlantic region, Climate Dynamics, doi: 10.1007/s00382-011-1077-2

[Compo et al. 2011] Compo G., Whitaker J., Sardeshmukh P., Matsui N., Allan R., Yin X., Gleason B., Vose R., Rutledge G., Bessemoulin P., Brönnimann S., Brunet M., Crouthamel R., Grant A., Groisman P., Jones P., Kruk M., Kruger A., Marshall G., Maugeri M., Mok H., Nordli Ø., Ross T., Trigo R., Wang X., Woodruff S., Worley S. (2011) The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1-28

[Demuzere et al. 2009] Demuzere M., Werner M., van Lipzig N., Roeckner E. (2009) An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns, International Journal of Climatology, 29, 1796-1810

[El-Kadi and Smithson 1992] El-Kadi A., Smithson P. (1992) Atmospheric classifications and synoptic climatology, Progress in Physical Geography, 16(4), 432-455

[Enke and Spekat 1997] Enke W., Spekat A. (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression, Climate Research, 8, 195-207

[Fettweis et al. 2011a] Fettweis X., Belleflamme A., Erpicum M., Franco B., Nicolay S. (2011a) Estimation of the sea-level rise by 2100 resulting from changes in the surface mass balance of the Greenland ice sheet. In: Climate Change - Geophysical Foundations and Ecological Effects, Blanco J., Kheradmand

H. (eds.). InTech, available on: http://www.intechopen.com/articles/show/title/estimation-of-the-sea-level-rise-by-2100-resulting-from-changes-in-the-surface-mass-balance-of-the-g

[Fettweis et al. 2011b] Fettweis X., Mabille G., Erpicum M., Nicolay S., Van den Broeke M. (2011b) The 1958-2009 Greenland ice sheet surface melt and the mid-tropospheric atmospheric circulation, Climate Dynamics, doi: 10.1007/s00382-010-0772-8

[Fettweis et al. 2012] Fettweis X., Lang C., Belleflamme A., Erpicum M., Gallée H. (2012) Important role of the mid-tropospheric atmospheric circulation in the recent surface melt increase over the Greenland ice sheet, Geophysical Research Letters, in preparation

[Franco et al. 2011] Franco B., Fettweis X., Erpicum M., Nicolay S. (2011) Present and Future Climates of the Greenland Ice Sheet according to the IPCC AR4 models, Climate Dynamics, 36, 1897-1918, doi: 10.1007/s00382-010-0779-1

[Gutmann et al. 2011] Gutmann E., Rasmussen R., Liu C., Ikeda K., Gochis D., Clark M., Dudhia J., Thompson G. (2011) A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain, Journal of Climate, doi: 10.1175/2011JCLI4109.1

[Hanna et al. 2008] Hanna E., Huybrechts P., Steffen K., Cappelen J., Huff R., Shuman C., Irvine-Fynn T., Wise S., Griffiths M. (2008) Increased runoff from melt from the Greenland ice sheet: a response to global warming, Journal of Climate, 21, 331-341

[Hanna et al. 2009] Hanna E., Cappelen J., Fettweis X., Huybrechts P., Luckman A., Ribergaard M.H. (2009) Hydrologic response of the Greenland ice sheet: the role of oceanographic warming, Hydrological Processes (Special Issue: Hydrologic Effects of a Shrinking Cryosphere), 23(1), 7-30, doi: 10.1002/hyp.7090

[Huth 2000] Huth R. (2000) A circulation classification scheme applicable in GCM studies, Theoretical and Applied Climatology, 67, 1-18

[Huth et al. 2007] Huth R., Ustrnul Z., Dittmann E., Bissolli P., Pasqui M., James P. (2007) Inventory of circulation classification methods and their applications in Europe within the COST 733 action, Proceedings from the 5th annual meeting of the European Meteorological Society, Session AW8 Weather types classifications, Utrecht, 9-15

[Kalkstein et al. 1987] Kalkstein L., Tan G., Skindlov J. (1987) An evaluation of three clustering procedures for use in synoptic climatological classification, Journal of Climate and Applied Meteorology, 26, 717-730

[Kalnay et al. 1996] Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Leetmaa A., Reynolds B., Chelliah M., Ebisuzaki W., Higgins W., Janowiak J., Mo K., Ropelewski C., Wang J., Jenne R., Joseph D. (1996) The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437-471

[Kirchhofer 1973] Kirchhofer W. (1973) Classification of European 500mb patterns. Arbeitsbericht der Schweizerischen Meteorologischen Zentralanstalt, 43p, Geneva

[Kysely and Huth 2006] Kysely J., Huth R. (2006) Changes in atmospheric circulation over Europe detected by objective and subjective methods, Theoretical and Applied Climatology, 85, 19-36

[Lund 1963] Lund I.A. (1963) Map-pattern classification by statistical methods, Journal of Applied Meteorology, 2, 56-65

[Masson and Knutti 2011] Masson D., Knutti R. (2011) Spatial-scale dependence of climate model performance in the CMIP3 ensemble, Journal of Climate, 24, 2680-2692

[Meehl et al. 2007] Meehl G., Stocker T., Collins W., Friedlingstein P., Gaye A., Gregory J., Kitoh A., Knutti R., Murphy J., Noda A., Raper S., Watterson I., Weaver A., Zhao Z.-C. (2007) Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt K., Tignor M., Miller H. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

[Moss et al. 2010] Moss R., Edmonds J., Hibbard K., Manning M., Rose S., van Vuuren D., Carter T., Emori S., Kainuma M., Kram T., Meehl G., Mitchell J., Nakicenovic N., Riahi K., Smith S., Stouffer R., Thomson A., Weyant J., Wilbanks T. (2010) The next generation of scenarios for climate change research and assessment, Nature, 463, doi:10.1038/nature08823

[Mote 1998a] Mote T.L. (1998a) Mid-tropospheric Circulation and Surface Melt on the Greenland Ice Sheet. Part I: Atmospheric Teleconnections, International Journal of Climatology, 18, 111-130

[Mote 1998b] Mote T.L. (1998b) Mid-tropospheric Circulation and Surface Melt on the Greenland Ice Sheet. Part II: Synoptic Climatology, International Journal of Climatology, 18, 131-146

[Overland et al. 2011] Overland J., Wang M., Bond N., Walsh J., Kattsov V., Chapman W. (2011) Considerations in the selection of Global Climate Models for regional climate projections: the Arctic as a case study, Journal of Climate, 24, 1583-1597

[Pastor and Casado 2012] Pastor M., Casado M. (2012) Use of circulation types classifications to evaluate AR4 climate models over the Euro-Atlantic region, Climate Dynamics, doi:10.1007/s00382-012-1449-2

[Philipp et al. 2007] Philipp A., Della-Marta P., Jacobeit J., Fereday D., Jones P., Moberg A., Wanner H. (2007) Long-term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering, Journal of Climate, 20, 4065-4095

[Philipp et al. 2010] Philipp A., Bartholy J., Beck C., Erpicum M., Esteban P., Fettweis X., Huth R., James P., Jourdain S., Kreienkamp F., Krennert T., Lykoudis S., Michalides S., Pianko K., Post P., Rassilla Álvarez D., Schiemann R., Spekat A., Tymvios F. S. (2010) COST733CAT - a database of weather and circulation type classifications, Physics and Chemistry of the Earth, 35(9-12), 360-373

[Plaut and Simonnet 2001] Plaut G., Simonnet E. (2001) Large-scale circulation classification, weather regimes, and local climate over France, the Alps and Western Europe, Climate Research, 17, 303-324

[Randall et al. 2007] Randall D., Wood R., Bony S., Colman R., Fichefet T., Fyfe J., Kattsov V., Pitman A., Shukla J., Srinivasan J., Stouffer R., Sumi A., Taylor K. (2007) Climate models and their evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt K., Tignor M., Miller H. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

[Reifen and Toumi 2009] Reifen C., Toumi R. (2009) Climate projections: Past performance no guarantee of future skill? Geophysical Research Letters, 36, doi:10.1029/2009GL038082

[Schuenemann and Cassano 2009] Schuenemann K., Cassano J. (2009) Changes in synoptic weather patterns and Greenland precipitation in the 20th and 21st centuries: 1. Evaluation of late 20th century simulations from IPCC models, Journal of Geophysical Research, 114, doi:10.1029/2009JD011705

[Schuenemann and Cassano 2010] Schuenemann K., Cassano J. (2010) Changes in synoptic weather patterns and Greenland precipitation in the 20th and 21st centuries: 2. Analysis of 21st century atmospheric changes using self-organizing maps, Journal of Geophysical Research, 115, doi:10.1029/2009JD011706

[Stoner et al. 2009] Stoner A. M., Hayhoe K., Wuebbles D. (2009) Assessing General Circulation Model simulations of atmospheric teleconnection patterns, Journal of Climate, 22, 4348-4372

[Tedesco et al. 2008] Tedesco M., Serreze M., Fettweis X. (2008) Diagnosing the extreme surface melt event over southwestern Greenland in 2007, The Cryosphere, 2, 159-166

[Uppala et al. 2005] Uppala, S. M., Kallberg P. W., Simmons A. J., Andrae U., da Costa Bechtold V., 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., van de Berg L., Bidlot J., Bormann N., Caires S., Chevallier F., Dethof A., Dragosavac M., Fisher M., Fuentes M., Hagemann S., Holm E., Hoskins B., 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., Woollen J. (2005) The ECMWF re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961-3012, doi:10.1256/qj.04.176

[Vautard and Yiou 2009] Vautard R., Yiou P. (2009) Control of recent European surface climate change by atmospheric flow, Geophysical Research Letters, 36, doi: 10.1029/2009GL040480

[Walsh et al. 2008] Walsh J., Chapman W., Romanovsky V., Christensen J., Stendel M. (2008) Global climate model performance over Alaska and Greenland, Journal of Climate, 21, 6156-6174

[Wilby and Wigley 2000] Wilby R., Wigley T. (2000) Precipitation predictors for downscaling: observed and general circulation model relationships, International Journal of Climatology, 20, 641-661

[Yarnal et al. 2001] Yarnal B., Comrie A., Frakes B., Brown D. (2001) Developments and prospects in synoptic climatology, International Journal of Climatology, 21(15), 1923-1950

[Yoshimori and Abe-Ouchi 2011] Yoshimori M., Abe-Ouchi A. (2011) Sources of spread in multi-model projections of the Greenland ice-sheet surface mass balance, Journal of Climate, doi: 10.1175/2011JCLI4011.1

[Zorita et al. 1995] Zorita E., Hughes J., Lettemaier D., von Storch H. (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, Journal of Climate, 8, 1023-1042

[Zorita and von Storch 1999] Zorita E., von Storch H. (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods, Journal of Climate, 12, 2474-2489

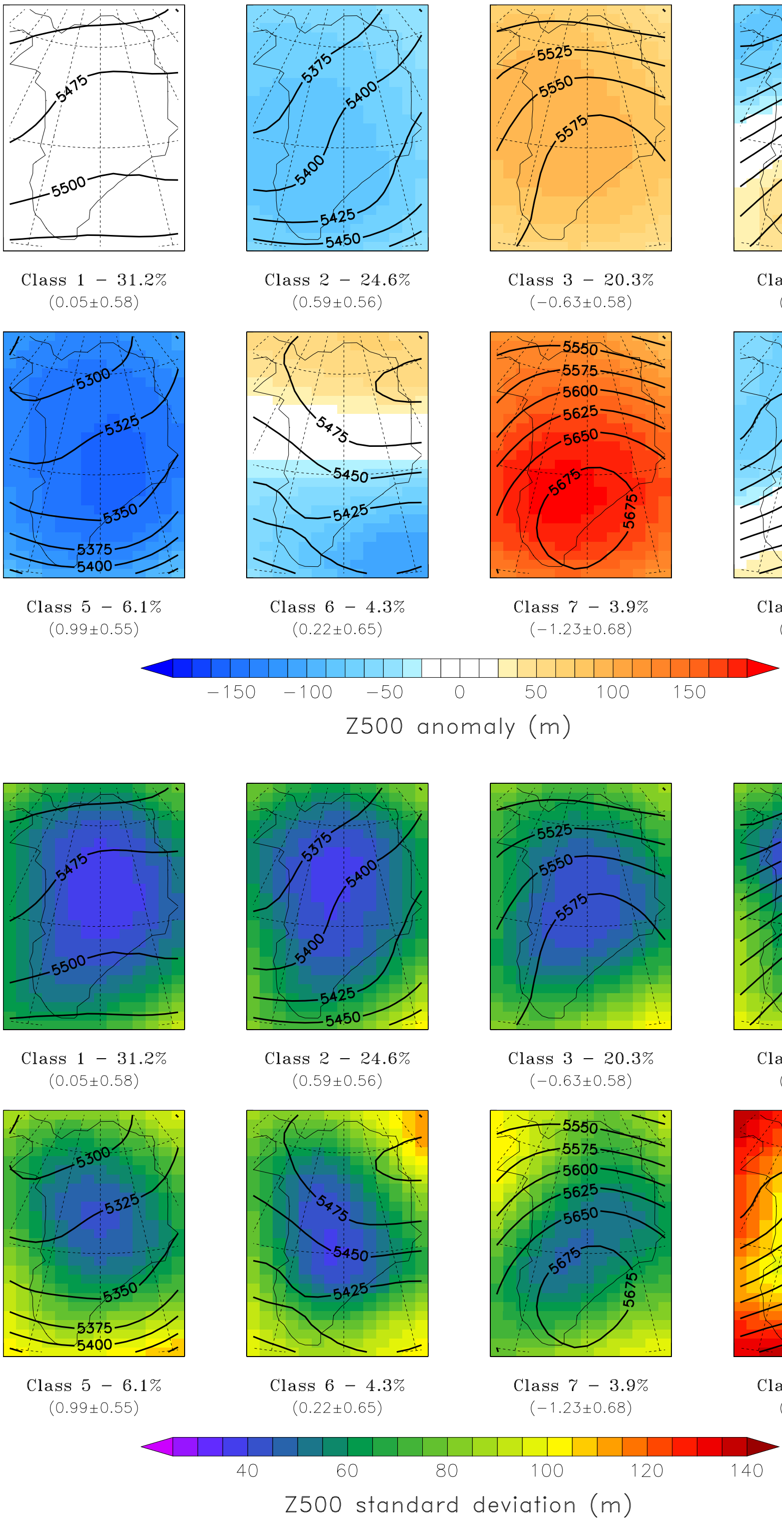

Fig. 1 The JJA circulation types during the period 1961-1990 for the automatic circulation type classification using the Euclidean distance for ERA-40 over Greenland are represented by the solid black isohypses (in metres). The relative frequency of each type is shown in bold and the mean CPC (Climate Prediction Center) NAO index of each class as well as its standard deviation are listed in brackets. Top: the anomaly is calculated as the difference between the class mean Z500 and $Z500_{JJA}$ from 1961-1990. Bottom: the colours represent the standard deviation of each class

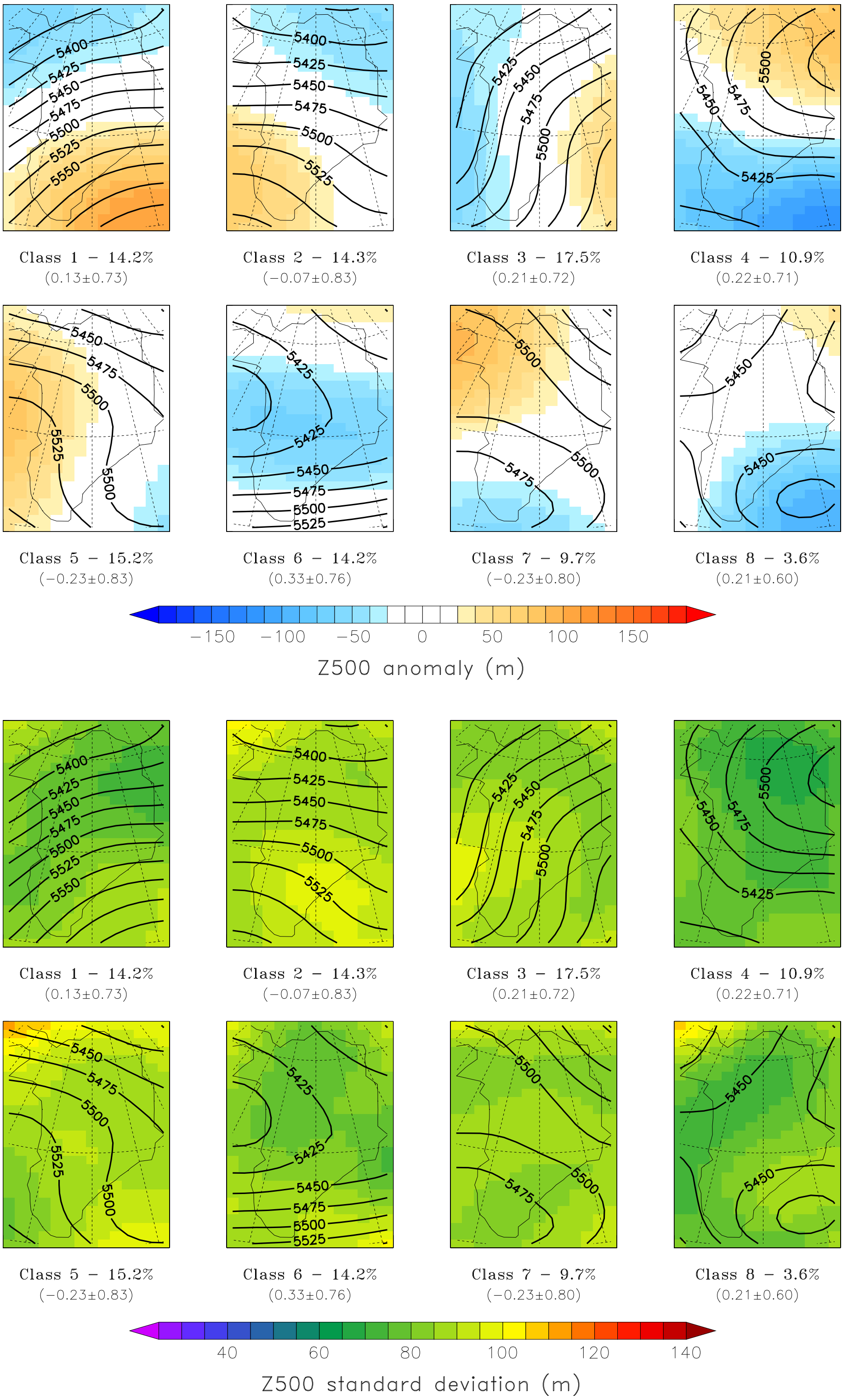

Fig. 2 The JJA circulation types from 1961-1990 for the automatic circulation type classification using the Spearman rank correlation for ERA-40 over Greenland are represented by the solid black isohypses (in metres). The relative frequency of each type is shown in bold and the mean CPC NAO index of each class as well as its standard deviation are listed in brackets. Top: the anomaly is calculated as the difference between the class mean Z500 and $Z500_{JJA}$ for the period 1961-1990. Bottom: the colours represent the standard deviation of each class

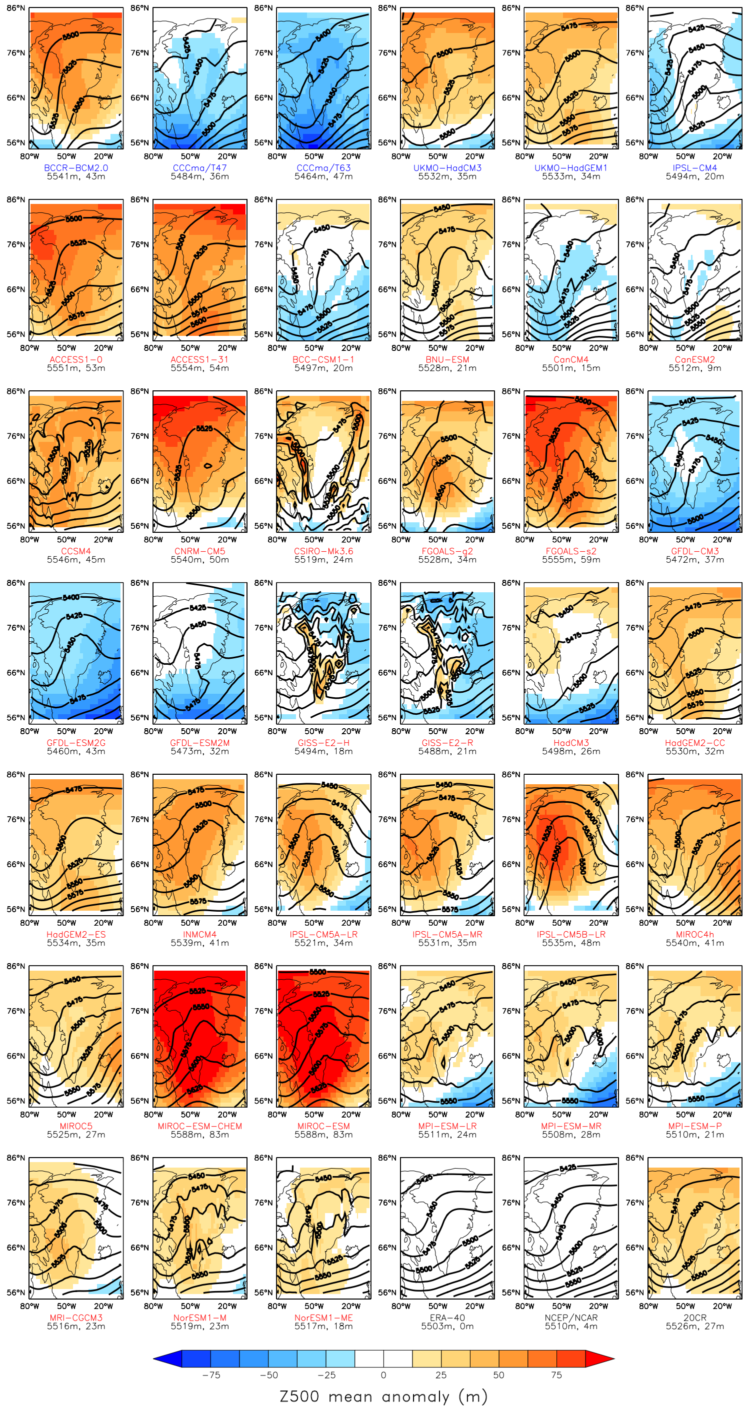

Fig. 3 The simulated $Z500_{JJA}$ from each GCM and reanalysis from 1961-1990 is represented by the black isohypses (in metres). The anomaly is calculated as the difference between the GCM/reanalysis $Z500_{JJA}$ (shown below each plot, on the left) and the ERA-40 $Z500_{JJA}$. The root mean square error between the GCM/reanalysis and the ERA-40 $Z500_{JJA}$ is also listed (below each plot, on the right). The CMIP3 GCMs are marked in blue, the CMIP5 GCMs in red and the reanalyses in black. GCMs for which only monthly data are available are shown to give an idea of the spread of $Z500_{JJA}$

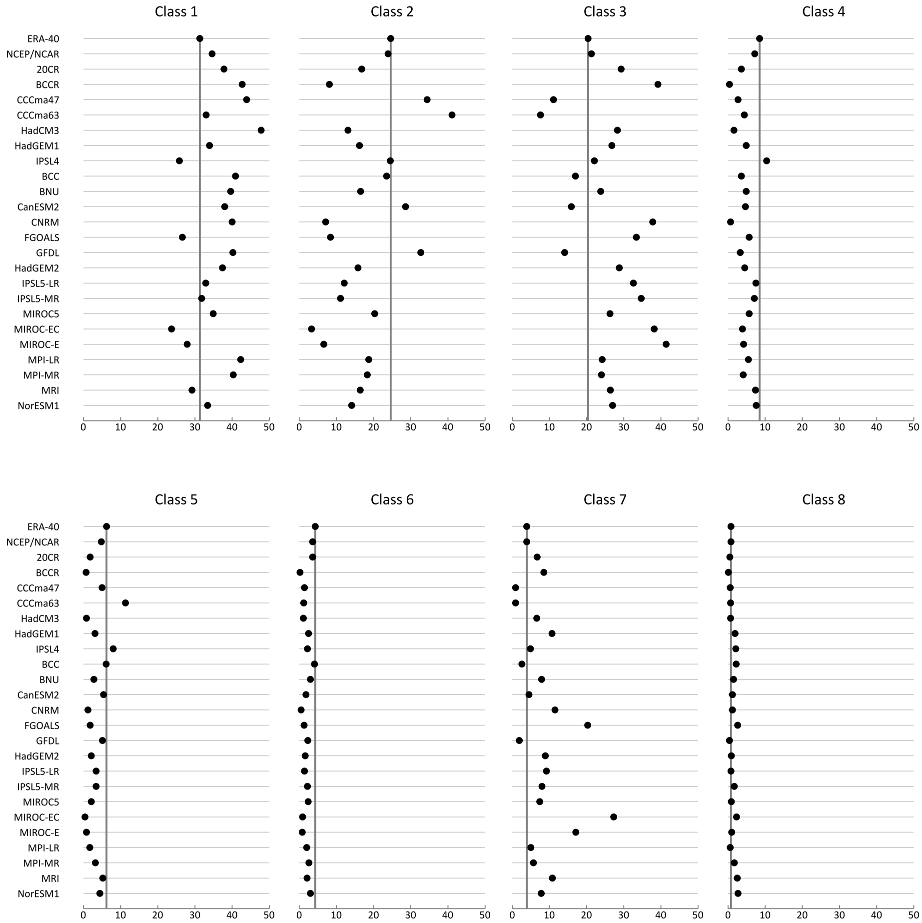

Fig. 4 The frequency (in %) of each circulation type of the Euclidean distance classification is represented for all GCMs and reanalyses for summer (JJA) for the period 1961-1990. The solid grey line is the ERA-40 frequency

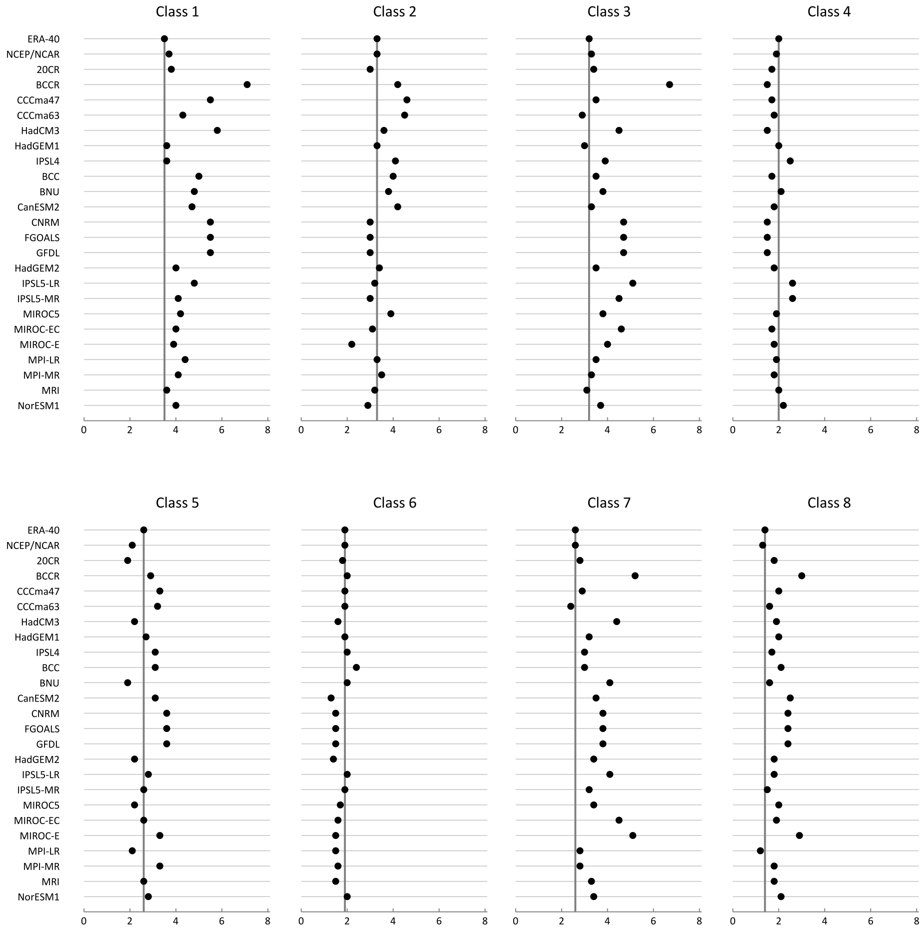

Fig. 5 The mean persistence (in days) of each circulation type of the Euclidean distance classification based on ERA-40 is represented for all GCMs and reanalyses for summer (JJA) for the period 1961-1990. The solid grey line is the ERA-40 persistence

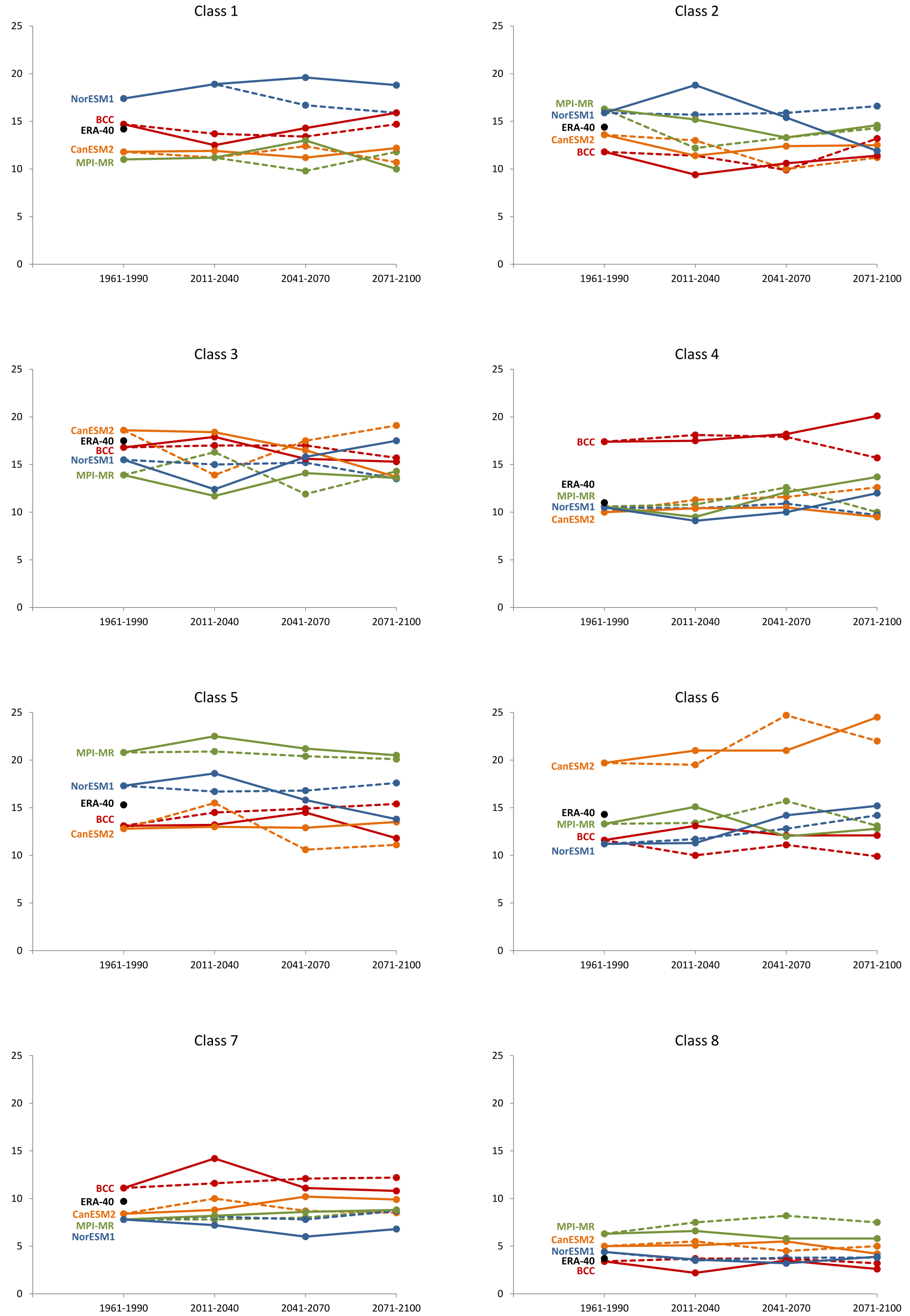

Fig. 6 The frequency (in %) of each circulation type is represented for the retained GCMs for the Spearman rank correlation classification for the Historical experiment and the three future periods for the RCP4.5 experiment (dashed line) and the RCP8.5 experiment (solid line). The ERA-40 frequency is shown for comparison

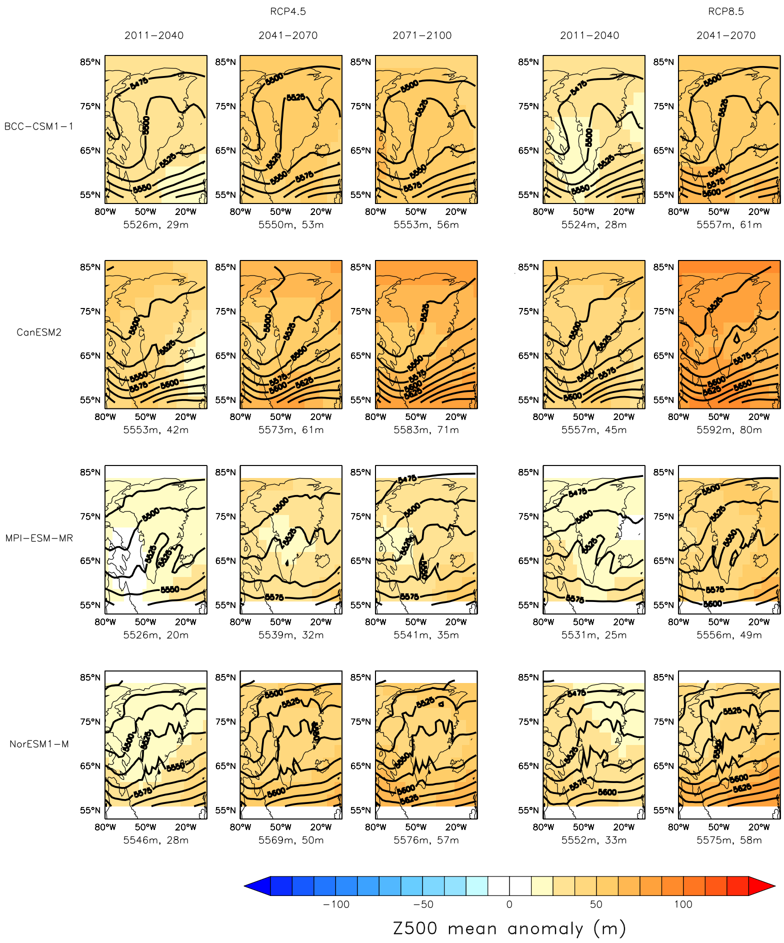

Fig. 7 The projected $Z500_{JJA}$ of some CMIP5 GCMs for the three future periods is represented by the black isohypses (in metres) for both future projection experiments. The anomaly is calculated as the difference between the GCM future period $Z500_{JJA}$ (shown below each plot, on the left) and its current climate (Historical experiment, 1961-1990) $Z500_{JJA}$. The root mean square error between the GCM future period and current climate $Z500_{JJA}$ is also listed below each plot, on the right

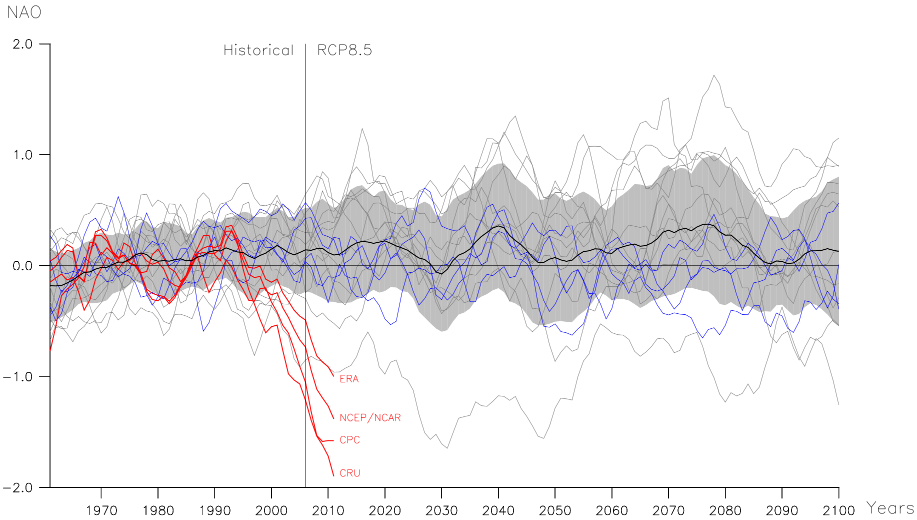

Fig. 8 The mean summer (JJA) NAO (North Atlantic Oscillation) index is normalized by 1961-1990 and shown as 10-year running mean. For the GCMs, the Historical experiment is plotted from 1961-2005 and the RCP8.5 from 2006-2100. The four GCMs used for the future projections are drawn in blue, the others in grey. The GCM mean is shown in black and the one standard deviation interval around this mean is shaded in grey. ERA is divided into ERA-40 from 1961-1999 and ERA-Interim from 2000-2011. The CRU (Climatic Research Unit, see http://www.cru.uea.ac.uk/cru/data/nao/ for more details) and CPC (Climate Prediction Center, see http://www.cpc.ncep.noaa.gov/ for more details) NAO indices are also shown

Table 1 A short name has only been assigned to the GCMs/reanalyses for which we could obtain daily data of the geopotential height at 500 hPa. The data were downloaded from the website indicated as footnote, except for the CMIP3 monthly data, which come from the CMIP3 server at https://esg.llnl.gov:8443/

Model name	Short name	Spatial resolution (lat, lon)	Research centre ID (Country)
NCEP/NCAR Reanalysis 1 [1]	NCEP/NCAR	$2.5^o \times 2.5^o$	NCEP-NCAR (United States)
ECMWF ERA-40 [2]	ERA-40	$1.125^o \times 1.125^o$	ECMWF (Europe)
20thC-ReanV2 [3]	20CR	$2.0^o \times 2.0^o$	NOAA ESRL/PSD (United States)
BCCR-BCM2.0 [4]	BCCR	$2.8^o \times 2.8^o$	BCCR (Norway)
CCCma-CGCM3.1/T47 [5]	CCCma47	$3.75^o \times 3.75^o$	CCCma (Canada)
CCCma-CGCM3.1/T63 [5]	CCCma63	$2.8^o \times 2.8^o$	CCCma (Canada)
IPSL-CM4_v1 [6]	IPSL4	$2.5^o \times 3.75^o$	IPSL (France)
UKMO-HadCM3 [7]	HadCM3	$2.5^o \times 3.75^o$	MOHC (United Kingdom)
UKMO-HadGEM1 [7]	HadGEM1	$1.25^o \times 1.875^o$	MOHC (United Kingdom)
ACCESS1.0 [8]		$1.25^o \times 1.875^o$	CSIRO-BOM (Australia)
ACCESS1.3 [8]		$1.25^o \times 1.875^o$	CSIRO-BOM (Australia)
BCC-CSM1-1 [8]	BCC	$2.8^o \times 2.8^o$	BCC (China)
BNU-ESM [8]	BNU	$2.8^o \times 2.8^o$	BNU (China)
CanESM2 [8]	CanESM2	$2.8^o \times 2.8^o$	CCCma (Canada)
CNRM-CM5 [8]	CNRM	$1.4^o \times 1.4^o$	CNRM-CERFACS (France)
CSIRO-Mk3.6 [8]		$1.875^o \times 1.875^o$	CSIRO-QCCCE (Australia)
FGOALS-s2 [8]	FGOALS	$1.67^o \times 2.8^o$	LASG-IAP(China)
GFDL-ESM2M [8]	GFDL	$2.0^o \times 2.5^o$	NOAA GFDL (United States)
GISS-E2-H [8]		$2.0^o \times 2.5^o$	NASA-GISS (United States)
GISS-E2-R [8]		$2.0^o \times 2.5^o$	NASA-GISS (United States)
HadCM3 [8]		$2.5^o \times 3.75^o$	MOHC (United Kingdom)
HadGEM2-CC [8]	HadGEM2	$1.25^o \times 1.875^o$	MOHC (United Kingdom)
HadGEM2-ES [8]		$1.25^o \times 1.875^o$	MOHC (United Kingdom)
INMCM4 [8]		$1.5^o \times 2.0^o$	INM (Russia)
IPSL-CM5A-LR [8]	IPSL5-LR	$1.875^o \times 3.75^o$	IPSL (France)
IPSL-CM5A-MR [8]	IPSL5-MR	$1.25^o \times 2.5^o$	IPSL (France)
IPSL-CM5B-LR [8]		$1.875^o \times 3.75^o$	IPSL (France)
MIROC4h [8]		$0.56^o \times 0.56^o$	MIROC (Japan)
MIROC5 [8]	MIROC5	$1.4^o \times 1.4^o$	MIROC (Japan)
MIROC-ESM-CHEM [8]	MIROC-EC	$2.8^o \times 2.8^o$	MIROC (Japan)
MIROC-ESM [8]	MIROC-E	$2.8^o \times 2.8^o$	MIROC (Japan)
MPI-ESM-LR [8]	MPI-LR	$1.875^o \times 1.875^o$	MPI-M (Germany)
MPI-ESM-MR [8]	MPI-MR	$1.875^o \times 1.875^o$	MPI-M (Germany)
MRI-CGCM3 [8]	MRI	$1.125^o \times 1.125^o$	MRI (Japan)
NorESM1-M [8]	NorESM1	$1.875^o \times 2.5^o$	NCC (Norway)

[1] http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml

[2] http://www.ecmwf.int/products/data/archive/descriptions/e4/index.html

[3] http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html

[4] https://esg.llnl.gov:8443/

[5] http://www.cccma.ec.gc.ca/data/data.shtml

[6] http://mc2.ipsl.jussieu.fr/simules.html

[7] http://badc.nerc.ac.uk/home/index.html

[8] http://pcmdi3.llnl.gov/esgcet

Table 2 The root mean square error (RMSE) is calculated over the frequency differences between the GCM/reanalysis and ERA-40 for the classifications using the Euclidean distance (DIST) and the Spearman rank correlation (RANK) for the current climate (1961-1990, JJA). The other columns indicate the number of classes that have a significantly different intraclass distribution (at 5%) with regard to ERA-40 and on the basis of a two-sample Kolmogorov-Smirnov test

	DIST		RANK	
	RMSE	KS-test	RMSE	KS-test
NCEP/NCAR	1.44	0	0.74	0
20CR	5.42	6	1.54	8
BCCR	10.56	5	3.47	8
CCCma47	7.02	5	3.26	7
CCCma63	7.89	5	2.55	8
HadCM3	8.41	5	2.07	8
HadGEM1	4.9	4	2.17	8
IPSL4	2.44	6	2.46	8
BCC	4.09	2	2.8	6
BNU	4.86	6	2.98	6
CanESM2	3.58	3	2.44	2
CNRM	10.28	5	3.56	8
FGOALS	9.76	7	3.05	8
GFDL	5.26	6	3.32	6
HadGEM2	5.63	5	2.44	8
IPSL5-LR	6.68	4	6.68	8
IPSL5-MR	7.28	4	5.09	7
MIROC5	3.67	5	1.2	8
MIROC-EC	13.45	7	3.88	8
MIROC-E	11.26	6	4.4	8
MPI-LR	5.1	1	2.75	6
MPI-MR	4.59	2	2.95	7
MRI	4.54	3	3.9	7
NorESM1	4.81	3	2.08	7

Table 3 The root mean square error (RMSE) is calculated over the frequency differences between the retained CMIP5 GCMs and ERA-40 (1961-1990) for the classifications using the Spearman rank correlation (upper part) and the Euclidean distance after removing the $Z500_{JJA}$ increase (lower part)

	Historical	RCP4.5			RCP8.5		
	1961-1990	2011-2040	2041-2070	2071-2100	2011-2040	2041-2070	2071-2100
BCC	2.8	3.21	3.24	2.58	3.5	3.12	3.86
CanESM2	2.44	2.63	4.4	3.67	2.93	2.93	4.06
MPI-MR	2.95	2.89	3.64	2.7	3.7	2.8	3.07
NorESM1	2.08	2.3	1.69	2.0	3.51	2.46	2.24
BCC	4.09	3.24	3.12	2.52	3.12	2.36	1.79
CanESM2	3.58	2.07	1.66	2.45	3.65	1.99	2.32
MPI-MR	4.59	4.46	4.19	4.47	4.63	4.27	3.53
NorESM1	4.81	4.97	4.57	4.86	4.24	3.97	4.77

Table 4 The mean persistence (in days) is shown for each circulation type for the last period (2071-2100, JJA) of the future experiments RCP4.5 (upper part) and RCP8.5 (lower part) using the Spearman rank correlation classification based on ERA-40

	1	2	3	4	5	6	7	8
BCC	2.9	2.1	2.4	3.4	2.4	1.9	2.2	1.3
CanESM2	2.2	1.9	2.2	2.9	2.0	2.3	1.8	1.4
MPI-MR	2.2	2.1	2.1	2.8	2.4	2.1	1.9	1.5
NorESM1	2.9	2.2	2.1	2.7	2.3	2.1	2.2	1.5
BCC	2.9	2.1	2.4	4.4	2.1	2.3	2.2	1.2
CanESM2	2.6	2.0	1.9	2.9	2.1	2.7	2.0	1.2
MPI-MR	2.3	2.1	2.1	3.1	2.5	2.0	2.0	1.4
NorESM1	3.2	2.1	2.4	3.0	2.3	2.1	1.9	1.3