A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns

Swen Brands[1], Juan Antonio Fernández-Granja[2], joaquin Bedia[3], Ana Casanueva[3], and Jesús Fernández[2]

[1]MeteoGalicia - Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia
[2]Instituto de Física de Cantabria, Universidad de Cantabria-CSIC
[3]University of Cantabria

March 9, 2023

Abstract

The performance of 61 global climate models participating in CMIP5 and 6 is evaluated for the Southern Hemisphere extratropics in terms of typical regional-scale atmospheric circulation patterns. These patterns are known to be linked with a number of key variables in atmospheric physics and chemistry and provide an overarching concept for model evaluation. First, hemispheric-wide error and ranking maps are provided for each model and regional details are described. Then, the results are compared with those obtained in a companion study for the Northern Hemisphere. For most models, the average error magnitude and ranking position is similar on both hemispheres, ruling out systematic tuning towards either of the two. CMIP6 models perform better on average than CMIP5 models and the interactive simulation of more climate system components does not deteriorate the results for most model families. Better performance is associated with higher resolution in the atmosphere, following a non-linear relationship.

A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns

S. Brands[1], J.A. Fernández-Granja[2], J. Bedía[3,4], A. Casanueva[3,4], J. Fernández[2]

[1]MeteoGalicia, Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia, 15707 Santiago de Compostela, Spain
[2]Instituto de Física de Cantabria, Universidad de Cantabria-CSIC, 39005 Santander, Spain
[3]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, 39005 Santander, Spain
[4]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, 39005 Santander, Spain

Key Points:

- CMIP6 models perform better than CMIP5 models on average
- Southern Hemisphere model ranking similar to Northern Hemisphere ranking
- More complex model versions perform similar to less complex ones

Corresponding author: Swen Brands, `swen.brands@gmail.com`

Abstract

The performance of 61 global climate models participating in CMIP5 and 6 is evaluated for the Southern Hemisphere extratropics in terms of typical regional-scale atmospheric circulation patterns. These patterns are known to be linked with a number of key variables in atmospheric physics and chemistry and provide an overarching concept for model evaluation. First, hemispheric-wide error and ranking maps are provided for each model and regional details are described. Then, the results are compared with those obtained in a companion study for the Northern Hemisphere. For most models, the average error magnitude and ranking position is similar on both hemispheres, ruling out systematic tuning towards either of the two. CMIP6 models perform better on average than CMIP5 models and the interactive simulation of more climate system components does not deteriorate the results for most model families. Better performance is associated with higher resolution in the atmosphere, following a non-linear relationship.

Plain Language Summary

This letter provides a survey on the capability of global climate models to reproduce the regional atmospheric circulation in the Southern Hemisphere in present climate conditions. Climate models from the latest model generation are found to perform better on average than those of the previous generation and the obtained model ranking is similar to that found for the Northern Hemisphere in a companion study. While model performance is found to be generally unrelated to model complexity in terms of covered climate system components, better results are associated with higher model resolution in the atmosphere.

1 Introduction

The vast ocean and ice-sheet areas in the Southern Hemisphere (SH) extratropics are virtually inhabited, but play a key role for the global climate system and are thus of paramount importance for mankind. In this context, the quasi-persistent circumpolar westerly winds blowing along the open sea channel in the mid-latitudes are of key relevance for several reasons. Partly offset by meso-scale ocean eddies tending to break up the intense ocean stratification, the westerlies drive the up-welling of carbon and nutrient-rich deep water and also force the Antarctic Circumpolar Current (ACC), which is embedded in the Meridional Overturning Circulation that in turn governs the low frequency variability of the global climate system (Abernathey et al., 2011; Speer & Marshall, 2012; Meredith et al., 2012; Hogg et al., 2017; Zhang et al., 2019). Commonly described by the Southern Annular Mode, also referred to as "Antarctic Oscillation" (Trenberth, 1979; Rogers & van Loon, 1982; Thompson & Wallace, 2000), the westerlies have shifted poleward during the last decades while, simultaneously, the Hadley Cell and associated large-scale subsidence in the sub-tropics have intensified (Thompson et al., 2000; Nguyen et al., 2015; Fogt & Marshall, 2020). Both anomalies are projected to magnify during the course of the 21^{st} century in global climate model (GCM) experiments (Deng et al., 2022), leading to more frequent extreme events like, e.g., droughts (Holgate et al., 2020) or sea-surface warming events (Duran et al., 2020) whose accumulated effects also alter the mass balance of the glaciers and ice-sheets in the SH.

While Patagonian glaciers are mainly affected by temperature and precipitation anomalies (Boex et al., 2013), melting into the Amundsen and Bellingshausen Seas is the main driver of Antarctic continental ice loss (Hughes, 1981; Rignot et al., 2019). The aforementioned poleward shift of the westerlies has led to an enhanced transport of relatively warm Circumpolar Deep Water, located at intermediate depths below the cold surface ocean layer, towards the continental shelf of the aforementioned sea areas (Steig et al., 2012), thereby thinning the ice shelves from below and melting the glaciers and ice streams at their ground lines. Subject to large uncertainties (Rignot et al., 2011), these processes contribute to global sea-level rise (Fox-Kemper et al., 2021).

Over glacial-to-interglacial cycles, the strength and position of the westerlies are also considered key to variations in the upwelling of carbon-rich Antarctic Bottom Water (AABW) reservoirs, associated with CO_2 degassing into the atmosphere (Sigman & Boyle, 2000; Speer & Marshall, 2012). There are indications that strong westerlies located near the Antarctic continent —well aligned with the ACC—, typically occur during warm, interglacial periods and enhance the aforementioned process leading to an increase in global CO_2 concentrations. Weaker westerlies located far away from the Antarctic continent and poorly aligned with the ACC are, in turn, currently discussed to be characteristic of cold, glacial periods. CO_2 degassing into the atmosphere wold be reduced in this case, favouring a net carbon storage in the AABW (Toggweiler et al., 2006; Gray et al., 2021). Finally, AABW formation itself is also controlled by wind forcing, namely by southerly katabatic winds blowing down the Antarctic continent, pushing the sea-ice away from coast and thereby forming coastal polynyas. In these ocean water areas surrounded by sea-ice, the nutrient-rich upwelled waters are subject to brine rejection during sea-ice formation that leads to increase in salinity. Sinking to the ocean bottom is the consequence, where "preformed" nutrients can thereby accumulate. AABW formation is particularly productive in the Weddell and Ross Seas and is subject to pronounced low-frequency variability (Ito & Follows, 2005; Hogg et al., 2017; Silvano et al., 2020).

These considerations show that the atmospheric circulation in the SH extratropics, even in confined and relatively small regions such as the aforementioned sea areas, are relevant for the entire climate system. Consequently, comprehensive GCMs now extensively used in climate research should perform well in this regard.

The present study evaluates a large multi-model ensemble from the Coupled Model Intercomparison Projects 5 and 6 (Taylor et al., 2012; Eyring et al., 2016) in terms of the models' capability to reproduce the climatological frequencies of typical and recurrent patterns of the regional atmospheric circulation in the SH extratropics. To this aim, the Lamb Weather Types method (LWT), also known as Jenkison-Collison circulation typing approach (Lamb, 1972; Jenkinson & Collison, 1977; Jones et al., 1993) has been recently extended for systematic use in the SH (Fernández-Granja et al., 2023) and is here applied to 61 GCMs from CMIP5 and 6, and to 3 distinct reference reanalyses. The circulation types obtained from this method are known to correlate with many key variables in atmospheric physics and chemistry and therefore constitute an overarching concept to describe regional-scale climate variability. The method is thus complementary to those operating on larger scales used in a previous study (Bracegirdle et al., 2020) and it is a direct answer to the downscaling community's claim for process-based GCM evaluation based on the regional atmospheric circulation (Maraun et al., 2017; Røste & Landgren, 2022), here tailored to the SH mid-to-high latitudes (Olson et al., 2016; Fita et al., 2017; Charles et al., 2020; Evans et al., 2021). Together with the respective assessment for the Northern Hemisphere (Brands, 2022a), this study completes the picture of GCM performance in terms of regional atmospheric circulation patterns in the extratropics. Possible model tuning issues to either of the two hemispheres are also discussed.

2 Data and Methods

The study relies on 6-hourly instantaneous sea-level pressure data from 61 different GCM configurations participating in CMIP5 and 6, all retrieved from the ESGF data portals. *Historical* experiments are evaluated and the considered ensemble members for each GCM are listed in the *get_historical_metadata.py* function available from Brands et al. (2022). It will be shown that the role of internal model variability does not lead to substantial changes in the results (see Section 3.1). Since several EC-Earth model versions were found to be favoured when evaluated against ECMWF reanalyses (Dee et al., 2011; Hersbach et al., 2020) in the companion study conduced over the Northern Hemisphere (Brands, 2022a), the Japanese 55-year reanalysis (JRA-55) is here used as main reference dataset for model evaluation (Kobayashi et al., 2015).

The Jenkinson-Collison circulation types constitute an objective classification method based on the subjective approach made by Lamb (1972). This technique, also known as "Lamb Weather Types" (Jones et al., 1993), groups an instantaneous SLP pattern centered at a given grid-box into 27 classes depending on the direction of the geostrophic flow (or lack thereof) and the sign and strength of the vorticity. In addition to the purely cyclonic and anticyclonic types, there are 8 "purely directional" types —one for each of the 8 main cardinal directions— and 16 hybrid types characterized by a predominant flow from one of these directions combined with either cyclonic or anticyclonic conditions. A detailed description of the LWT method, including the extension to the SH relevant here, is provided in Fernández-Granja et al. (2023). The corresponding Python code is available from Brands et al. (2022). The LWT method is here applied in a rolling manner (Otero et al., 2017) looping through all boxes of a regular latitude-longitude grid covering a spatial domain extending from 30°S to 70°S with a 2.5° resolution. The considered time period is 1979 to 2005, for which data is available for all applied GCMs and reanalyses.

The LWT method is here said to be applicable for a given region if at least 20 types occur with a minimum relative frequency of 0.1% (i.e. n = 39 occurrences for 27 years and 6-hourly data) at the corresponding grid-box in the reference reanalysis (JRA-55). This criterion is fulfilled in virtually the entire study area. Moreover, to ensure that the regional-scale GCM ranking presented here is robust to a switch in the underlying reference reanalysis, ERA-Interim is evaluated against JRA-55 just as if it was another GCM and the obtained error is compared to the errors of the 61 authentic GCMs. If any of the considered GCMs is found to perform better than ERA-Interim at a given grid-box, this indicates large observational uncertainties in the corresponding region, leading to an exclusion of the grid-box from further assessment. Figure 1 shows that all grid-boxes over the Antarctic continent and adjacent sea-areas seasonally covered by sea-ice have to be excluded for this reason.

For consistency with Brands (2022a) and Brands (2022b), the mean absolute error (MAE) of the climatological relative frequencies of the $n = 27$ LWTs is used as main error measure at the grid-box scale:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|m_i - o_i| \tag{1}$$

where m_i and o_i are the relative frequencies of the i^{th} LWT (with $i = 1, 2, ..., 27$) from the GCM and reference reanalysis, respectively. Alternative error measures such as the *Transition Probability Matrix Score* (Fernandez-Granja et al., 2021) have been used as well, obtaining nearly identical results for the model ranking.

To explore the role of internal model variability due to initial conditions uncertainties, up to 18 distinct historical runs per GCM are evaluated for a subset of 13 GCMs, specified in Supplementary Figure 3. Then, the GCM complexity score proposed in Brands (2022a) is used, which is proportional to the number of Earth system components taken into account in the GCM and gives more weight on simulated than on prescribed components. This score is put into relation with the spatial median model performance over the SH in order to explore whether the more complex GCMs perform better or worse than the less complex ones on average. Finally, the SH results are plotted against the NH results obtained from Brands (2022a) in order to detect possible tuning efforts to either of the two hemispheres. For a comparison on equal terms, the NH results were modified by also removing the regions prone to substantial reanalysis uncertainties, as described above.

3 Results

3.1 Regional details

Figures 1 and 2 show the GCM ranking patterns based on the MAE for the 61 considered GCMs, with lower MAE values leading to better ranks and vice versa. The MAE values themselves are depicted in Supplementary Figures 1 and 2. Hereafter, individual GCMs will be grouped according to the atmosphere general circulation model (AGCM) used therein (Brands, 2022b).

From these figures, it can be seen that the IFS AGCM family (i.e. all EC-Earth GCM versions) performs best overall, followed by the HadGAM/UM AGCM family comprising the ACCESS and HadGEM GCMs, as well as KACE1.0G. All members of the HadGAM/UM family except ACCESS1.0 and HadGEM3-GC31-MM have similar error patterns, with larger errors in the Southern Ocean to the south, southwest and southeast of the Australian continent. ACCESS1.3, ACCESS-CM2, HadGEM3-GC31-MM and KACE1.0-G perform relatively poorly to the south and southwest of Cape of Good Hope and KACE1.0-G additionally performs badly off the east coast of South America. A large performance gain is observed from CSIRO-MK3.6 to the ACCESS GCM family, i.e. from the former to the present GCM family developed by CSIRO (see also Figure 3).

The GAMIL AGCM family comprising the FGOALS-g2 and g3 GCMs performs overly poorest in this multi-model comparison. The ECHAM AGCM family, including all MPI-ESM versions, AWI-ESM-1-1-LR, NESM3 and CMCC-CM, performs slightly worse than the IFS and HadGAM/UM families, except for MPI-ESM1-2-HR and MPI-ESM-1-2-HAM performing almost equally well. A particularly poor model performance is observed for NESM3 along virtually the entire subtropics, extending to the mid-latitudes in the South Atlantic Ocean, and for AWI-ESM-1-1-LR over the eastern South Pacific and eastern South Atlantic. The CAM AGCM family comprises the largest number of GCMs (CMCC-CM2-SR5, CMCC-CM2-HR4, CMCC-ESM2, CCSM4, NorESM1-M, NorESM2-LM, NorESM2-MM, SAM0-UNICON, TaiESM1, BCC-CSM1.1 and BCC-CSM2-MR) and yields intermediate to unfavourable ranks in most regions. CMCC-CM, NorESM2-MM and SAM0-UNICON perform best in this family, yielding very good ranks in specific regions. CanESM2 comprises the CanAM4 AGCM that is not used in any other GCM and performs relatively poorly.

For the ARPEGE AGCM family shown in Figure 2 (CNRM-CM5, CNRM-CM6-1, CNRM-CM6-1-HR and CNRM-ESM2-1), the model versions used in CMIP6 (CNRM-CM6-1, CNRM-CM6-1-HR and CNRM-ESM2-1) perform worse than the well performing CMIP5 version CNRM-CM5. Surprisingly, this decrease in model performance is particularly pronounced in the high resolution (HR) version. IFS (EC-Earth2.3, EC-Earth3, EC-Earth3-Veg, EC-Earth3-Veg-LR, EC-Earth3-AerChem and EC-Earth3-CC) is the best performing model family in the present study, obtaining very good ranks over a large fraction of the domain. Model ranks worse than 40 are very rare, except for the ocean area to the south of Africa in EC-Earth-Veg-LR. The performance of the GFDL-AM AGCM family comprising GFDL-CM3, GFDL-CM4, GFDL-ESM2G, GFDL-ESM4 and KIOST-ESM is similar in magnitude to the ECHAM family, with best results overall for GFDL-ESM4. In case of the GISS-E2 AGCM family, the use of the Russel ocean model in GISS-E2-R leads to substantially better results than the use of the HYCOM model used in GISS-E2-H, the configuration of these two GCMs being otherwise equal (Schmidt et al., 2014), and a further performance increase is obtained by the CMIP6 version GISS-E2.1-G, comparable to that obtained for the ECHAM and GFDL-AM families mentioned above (see Figure 3). The most pronounced performance gain from CMIP5 to 6 is obtained for the LMDZ AGCM family (i.e. from IPSL-CM5A-LR and MR to IPSL-CM6A-LR), yielding a MAE level for IPSL-CM6A-LR comparable to that obtained for the ECHAM and GFDL-AM families. The MIROC-AM family is prone to very large performance differences from the better performing versions MIROC5 and 6 to the substantially worse performing versions MIROC-ESM and MIROC-ES2L, which are both more complex (see Figure 4d). The performance of the GSMUV family decreases substantially from CMIP5 to 6 (from MRI-ESM1 and MRI-ESM2.0) whereas that of the INM-AM family increases drastically (from INM-CM4 to INM-CM5). Finally, IITM-ESM is one of the worst performing GCMs considered here, with large differences in the results from one region to another.

3.2 Performance Summary and Comparison with the NH results

In Figure 3, the hemispheric-wide MAE samples mapped in Supplementary Figures 1 and 2 are displayed in a single boxplot. Each item describes the error distribution of a specific GCM in terms of the median (horizontal black line), interquartile range (IQR, box) and whiskers extending to the full range, except for the outliers lying beyond $1.5 \times IQR$ below and above the 2^{nd} and 3^{rd} quartile, respectively. The last four boxes, depicted in light green, are built upon the joint samples of the more and the less complex GCMs used in CMIP5 and 6, respectively (outliers are not shown for these samples). To this end, the GCMs are grouped according to their complexity score obtained from Brands (2022a) and those obtaining a score ≥ 14 are considered more complex.

For both complexity classes, the models used in CMIP6 perform better on average than those used in CMIP5. The largest performance gains from CMIP5 to 6 are obtained for the FGOALS and IPSL GCMs. However, a performance loss is obtained for 4 GCM groups —ACCESS, CMCC-CM, CNRM-CM and MRI-ESM—, which are 2 groups more than for the NH results obtained in Brands (2022a). Supplementary Figure 3 shows that internal model variability does not substantially change the aforementioned results.

A comparison between the areal median performance in the SH vs. NH is provided in Figure 4a and b. Overall, GCM performance is better in the SH than in the NH (panel a), which may be simply due to the fact that GCMs tend to perform better over the ocean than over land (Brands, 2022a), the ocean area being much larger in the SH. A close correspondence is obtained for the median error samples of the two hemispheres, particularly if they are log-transformed and standardized separately in order to remove systematic differences in their hemisphere-specific shape, magnitude and dispersion (panel b). Largest deviations from the diagonal are obtained for CNRM-CM6-1-HR, MRI-ESM2, CMCC-CM2-SR5, CMCC-ESM2 and HadGEM3-GC31-MM, performing better over the NH, and for CSIRO-MK3.6, KIOST-ESM, GISS-E2-1-G, MPI-ESM1.2-HAM and INM-CM5, performing better over the SH.

A significant non-linear relationship is obtained between the resolution of the AGCM —here described by the number of grid-points constituting the global 3-dimensional mesh (longitudes $\times$ latitudes $\times$ vertical layers)—, and the median model performance, obtaining a Spearman correlation coefficient (rs) of -0.49. Higher resolution is associated with better performance, particularly above a threshold of approximately $1.8{\times}10^7$ grid-boxes. Note that CNRM-CM6-1-HR and CNRM-ESM2-1 are not shown in Figure 4c because they are out-of-scale due to their very high resolution. Interestingly, the corresponding link with the 3D resolution of the ocean sub-model is weak (rs = -0.29), yet significant at a test level of 5%.

Finally, median model performance over the SH is generally not associated with model complexity (r = -0.01) and, for most model families, the more complex versions perform at least equally well than the less complex ones (see Figure 4d). The MIROC-AGCM family is an exception in this sense, since the more complex model versions MIROC-ESM and MIROC-ES2L, probably due to their low horizontal resolution in the atmosphere (T42) (Brands et al., 2022), perform substantially worse than the less complex versions MIROC5 and 6 (T85).

4 Conclusions

In the present study, 61 different GCMs from CMIP5 and 6 have been evaluated in the SH extratropics excluding Antarctica, focusing on the models' ability to reproduce the climatological frequency of the 27 Lamb Weather Types, known to be associated with many environmental variables and thus constituting a overarching concept to regional-scale climate variability.

While all of the model families performing poorly in CMIP5 have improved considerably in CMIP6, most of the families already performing well in CMIP5 have suffered a slight performance loss. For most model families, the spatial average performance for the SH is similar to that obtained for the NH (Brands, 2022a), suggesting that systematic model tuning to either of the two hemispheres can be ruled out in general terms. For a small number of specific GCMs, however, substantial performance differences are obtained from one hemisphere to another and the reasons for this should be assessed in future studies. Whereas a higher resolution in the atmospheric sub-model of the considered GCMs is found to be associated with better performance, following an exponentially decreasing relationship, GCM complexity as defined in Brands (2022a) is generally unrelated to performance, except for the MIROC-AGCM family, whose more complex versions perform worse than the less complex ones over the SH. This is a promising result since the more complex models are also prone to more error sources. It is also an argument for the use of the more complex models, as they provide a more complete picture of the feedback processes governing the climate system (Séférian et al., 2019; Dunne et al., 2020; Döscher et al., 2021).

Open Research Section

Supplementary Figures 1 to 3 are contained in the Supporting Information (SI) file to this article, available from GRL's homepage. The Python source code underlying this study and the GCM metadata archive *get_historitcal_metadata.py* are publicly available from Brands et al. (2022) and so is the LWT dataset for the considered GCMs and reanalyses, retrievable from Brands et al. (2023b). Additional auxiliary material containing 1) separate pdf files for each error and ranking map, 2) netCDF files containing gridbox-scale GCM errors and 3) summary csv files listing the model complexity score from Brands (2022a) as well as the spatial median performance over the SH domain for each GCM can be retrieved from Brands et al. (2023a).

Acknowledgments

The authors acknowledge the public availability of the CMIP datasets via the ESGF data portals, as well the free distribution of the ECMWF and JMA reanalysis products. This study is part of the I+D+i project CORDyS (PID2020-116595RB-I00), funded by MCIN/AEI/10.13039/501100011033. J.A.F. acknowledges support from project ATLAS (PID2019-111481RB-I00), grant PRE2020-094728 funded by MCIN/AEI/10.13039/501100011033 and ESF investing in your future. A.C. acknowledges support from Project COMPOUND (TED2021-131334A-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. J.B. thanks his parents for their parental care in the early years of his life. S.B. would like to thank CESGA and AMTEGA for providing computational resources.

References

Abernathey, R., Marshall, J., & Ferreira, D. (2011). The dependence of Southern Ocean meridional overturning on wind stress. *Journal of Physical Oceanography*, *41*(12), 2261 - 2278. doi: 10.1175/JPO-D-11-023.1

Boex, J., Fogwill, C., Harrison, S., Glasser, N., Hein, A., Schnabel, C., & Xu, S. (2013). Rapid thinning of the Late Pleistocene Patagonian Ice Sheet followed migration of the Southern Westerlies. *Scientific reports*, *3*, 2118. doi: 10.1038/srep02118

Bracegirdle, T. J., Holmes, C. R., Hosking, J. S., Marshall, G. J., Osman, M., Patterson, M., & Rackow, T. (2020). Improvements in circumpolar Southern Hemisphere extratropical atmospheric circulation in CMIP6 compared to CMIP5. *Earth and Space Science*, *7*(6), e2019EA001065. doi: https://doi.org/10.1029/2019EA001065

Brands, S. (2022a). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. doi: https://doi.org/10.5194/gmd-15-1375-2022

Brands, S. (2022b). Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble. *Geophysical Research Letters*, *49*(23), e2022GL101446. doi: https://doi.org/10.1029/2022GL101446

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023a). Auxiliary online material to Brands et al. (2023): A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns. *figshare*. doi: 10.6084/m9.figshare.22193443.v1

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023b). Southern Hemisphere Lamb Weather Types from historical GCM experiments and various reanalyses (1.0) [data set]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.7612988

Brands, S., Tatebe, H., Danek, C., Fernández, J., Swart, N., Volodin, E., ... Tongwen, W. (2022). Python code to calculate Lamb circulation types for the Northern Hemisphere derived from historical CMIP simulations and reanalysis data [code]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.4555367

Charles, S. P., Chiew, F. H. S., Potter, N. J., Zheng, H., Fu, G., & Zhang, L. (2020). Impact of downscaled rainfall biases on projected runoff changes. *Hydrology and Earth System Sciences*, *24*(6), 2981–2997. doi: 10.5194/hess-24-2981-2020

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, *137*(656, Part a), 553-597. doi: 10.1002/qj.828

Deng, K., Azorin-Molina, C., Yang, S., Hu, C., Zhang, G., Minola, L., & Chen, D. (2022). Changes of Southern Hemisphere westerlies in the future warming climate. *Atmospheric Research*, *270*, 106040. doi: 10.1016/j.atmosres.2022.106040

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., ... Zhang, Q. (2021). The EC-Earth3 Earth System Model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development Discussions*, *2021*, 1–90. doi: 10.5194/gmd-2020-446

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., ... Zhao, M. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. doi: https://doi.org/10.1029/2019MS002015

Duran, E. R., England, M. H., & Spence, P. (2020). Surface ocean warming around Australia driven by interannual variability and long-term trends in Southern Hemisphere westerlies. *Geophysical Research Letters*, *47*(9), e2019GL086605. doi: https://doi.org/10.1029/2019GL086605

Evans, J., Virgilio, G., Hirsch, A., Hoffmann, P., Remedio, A. R., Ji, F., ... Coppola, E. (2021). The CORDEX-Australasia ensemble: evaluation and future projections. *Climate Dynamics*, *57*, 1385–1401. doi: 10.1007/s00382-020-05459-0

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: https://doi.org/10.5194/gmd-9-1937-2016

Fernandez-Granja, J. A., Casanueva, A., Bedia, J., & Fernández, J. (2021). Improved atmospheric circulation over Europe by the new generation of CMIP6 Earth System Models. *Climate Dynamics*, *56*, 3527–3540. doi: https://doi.org/10.1007/s00382-021-05652-9

Fernández-Granja, J. A., Brands, S., Bedia, J., Casanueva, A., & Fernández, J. (2023). Exploring the limits of the Jenkinson–Collison weather types classification scheme: a global assessment based on various reanalyses. *Climate Dynamics*. doi: 10.1007/s00382-022-06658-7

Fita, L., Evans, J., Argueso, D., King, A., & Liu, Y. (2017). Evaluation of the regional climate response in Australia to large-scale climate modes in the historical NARCliM simulations. *Climate Dynamics*, *49*, 1-15. doi: 10.1007/s00382-016-3484-x

Fogt, R. L., & Marshall, G. J. (2020). The Southern Annular Mode: Variability, trends, and climate impacts across the Southern Hemisphere. *WIREs Climate Change*, *11*(4), e652. doi: https://doi.org/10.1002/wcc.652

Fox-Kemper, B., Hewitt, H., Xiao, C., Aoalgeirsdottir, G., Drijfhout, S. S., Edwards, T., ... Yu, Y. (2021). Climate Change 2021: The Physical Science Basis. contribution of Working Group i to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In V. Masson-Delmotte et al. (Eds.), (pp. 1211–1362). Cambridge University Press. doi: 10.1017/9781009157896.011

Gray, W., de Lavergne, C., Jnglin Wills, R., Menviel, L., Spence, P., Holzer, M., … Michel, E. (2021). Poleward shift in the Southern Hemisphere westerly winds synchronous with the deglacial rise in CO2. *Research Square*, 1–50. doi: 10.21203/rs.3.rs-404786/v1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hogg, A. M., Spence, P., Saenko, O. A., & Downes, S. M. (2017). The energetics of Southern Ocean upwelling. *Journal of Physical Oceanography*, *47*(1), 135 - 153. doi: 10.1175/JPO-D-16-0176.1

Holgate, C. M., Van Dijk, A. I. J. M., Evans, J. P., & Pitman, A. J. (2020). Local and remote drivers of southeast Australian drought. *Geophysical Research Letters*, *47*(18), e2020GL090238. doi: https://doi.org/10.1029/2020GL090238

Hughes, T. J. (1981). The weak underbelly of the West Antarctic ice sheet. *Journal of Glaciology*, *27*(97), 518–525. doi: 10.3189/S002214300001159X

Ito, T., & Follows, M. (2005). Preformed phosphate, soft tissue pump and atmospheric CO2. *Journal of Marine Research*, *63*, 813–839. doi: 10.1357/0022240054663231

Jenkinson, A., & Collison, F. (1977). *An initial climatology of gales over the North Sea* (Vol. 62; Tech. Rep. No. 18). Bracknell, UK: Meterololgical Office.

Jones, P. D., Hulme, M., & Briffa, K. R. (1993). A comparison of Lamb circulation types with an objective classification scheme. *International Journal of Climatology*, *13*(6), 655-663. doi: https://doi.org/10.1002/joc.3370130606

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 Reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5-48. doi: 10.2151/jmsj.2015-001

Lamb, H. (1972). British Isles weather types and a register of daily sequence of circulation patterns, 1861-1971. *Geophysical Memoir*, *116*, 85pp. (HMSO)

Maraun, D., Shepherd, T., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J., … Mearns, L. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*, 764-773. doi: https://doi.org/10.1038/nclimate3418

Meredith, M. P., Garabato, A. C. N., Hogg, A. M., & Farneti, R. (2012). Sensitivity of the overturning circulation in the Southern Ocean to decadal changes in wind forcing. *Journal of Climate*, *25*(1), 99 – 110. doi: 10.1175/2011JCLI4204.1

Nguyen, H., Lucas, C., Evans, A., Timbal, B., & Hanson, L. (2015). Expansion of the Southern Hemisphere Hadley Cell in response to greenhouse gas forcing. *Journal of Climate*, *28*(20), 8067 - 8077. doi: 10.1175/JCLI-D-15-0139.1

Olson, R., Evans, J., Di Luca, A., & Argueso, D. (2016). The NARCliM project: Model agreement and significance of climate projections. *Climate Research*, *69*, 209-227. doi: 10.3354/cr01403

Otero, N., Sillmann, J., & Butler, T. (2017). Assessment of an extended version of the Jenkinson-Collison classification on CMIP5 models over europe. *Climate Dynamics*, 1559–1579. doi: 10.1007/s00382-017-3705-y

Rignot, E., Mouginot, J., Scheuchl, B., van den Broeke, M., van Wessem, M. J., & Morlighem, M. (2019). Four decades of Antarctic Ice Sheet mass balance from 1979-2017. *Proceedings of the National Academy of Sciences*, *116*(4), 1095-1103. doi: 10.1073/pnas.1812883116

Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A., & Lenaerts, J. T. M. (2011). Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophysical Research Letters*, *38*(5), L05503. doi: 10.1029/2011GL046583

Rogers, J. C., & van Loon, H. (1982). Spatial variability of sea level pressure and 500 mb height anomalies over the Southern Hemisphere. *Monthly Weather Review*, *110*(10), 1375 - 1392. doi: 10.1175/1520-0493(1982)110⟨1375:SVOSLP⟩2.0.CO;2

Røste, J., & Landgren, O. (2022). Impacts of dynamical downscaling on circulation type statistics in the Euro-CORDEX ensemble. *Climate Dynamics*, *59*, 2445–2466. doi: 10.1007/s00382-022-06219-y

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., … Zhang, J. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*(1), 141-184. doi: 10.1002/2013MS000265

Sigman, D., & Boyle, E. (2000). Glacial/interglacial variations in atmospheric carbon dioxide. *Nature*, *407*, 859-69. doi: 10.1038/35038000

Silvano, A., Foppert, A., Rintoul, S., Holland, P., Tamura, T., Kimura, N., … Macdonald, A. (2020). Recent recovery of Antarctic Bottom Water formation in the Ross Sea driven by climate anomalies. *Nature Geoscience*, *13*, 1-7. doi: 10.1038/s41561-020-00655-3

Speer, K., & Marshall, J. (2012). Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geoscience - NAT GEOSCI*, *5*, 171-180. doi: 10.1038/ngeo1391

Steig, E., Ding, Q., Battisti, D., & Jenkins, A. (2012). Tropical forcing of Circumpolar Deep Water Inflow and outlet glacier thinning in the Amundsen Sea Embayment, West Antarctica. *Annals of Glaciology*, *53*(60), 19–28. doi: 10.3189/2012AoG60A110

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., … Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182-4227. doi: 10.1029/2019MS001791

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485-498. doi: https://doi.org/10.1175/BAMS-D-11-00094.1

Thompson, D. W. J., & Wallace, J. M. (2000). Annular modes in the extratropical circulation. part I: Month-to-month variability. *Journal of Climate*, *13*(5), 1000 - 1016. doi: 10.1175/1520-0442(2000)013⟨1000:AMITEC⟩2.0.CO;2

Thompson, D. W. J., Wallace, J. M., & Hegerl, G. C. (2000). Annular modes in the extratropical circulation. part II: Trends. *Journal of Climate*, *13*(5), 1018 - 1036. doi: 10.1175/1520-0442(2000)013⟨1018:AMITEC⟩2.0.CO;2

Toggweiler, J. R., Russell, J. L., & Carson, S. R. (2006). Midlatitude westerlies, atmospheric CO2, and climate change during the ice ages. *Paleoceanography*, *21*(2). doi: https://doi.org/10.1029/2005PA001154

Trenberth, K. E. (1979). Interannual variability of the 500 mb zonal mean flow in the Southern Hemisphere. *Monthly Weather Review*, *107*(11), 1515 - 1524. doi: 10.1175/1520-0493(1979)107⟨1515:IVOTMZ⟩2.0.CO;2

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., … Little, C. M. (2019). A review of the role of the Atlantic Meridional Overturning Circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*(2), 316-375. doi: 10.1029/2019RG000644

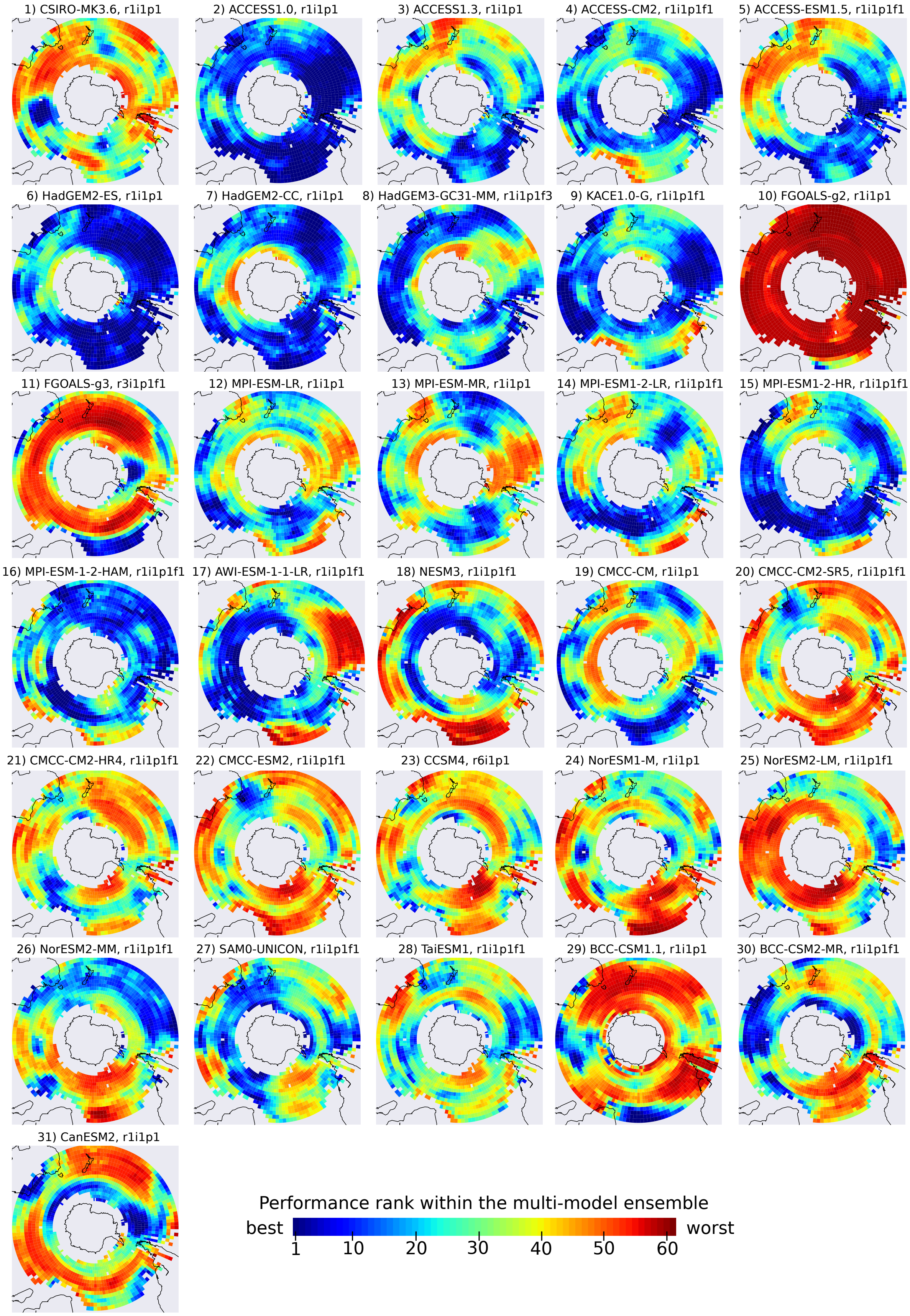

Figure 1. Ranking of the GCMs according to MAE defined in Equation 1, reference: JRA-55, 1979-2005, part 1

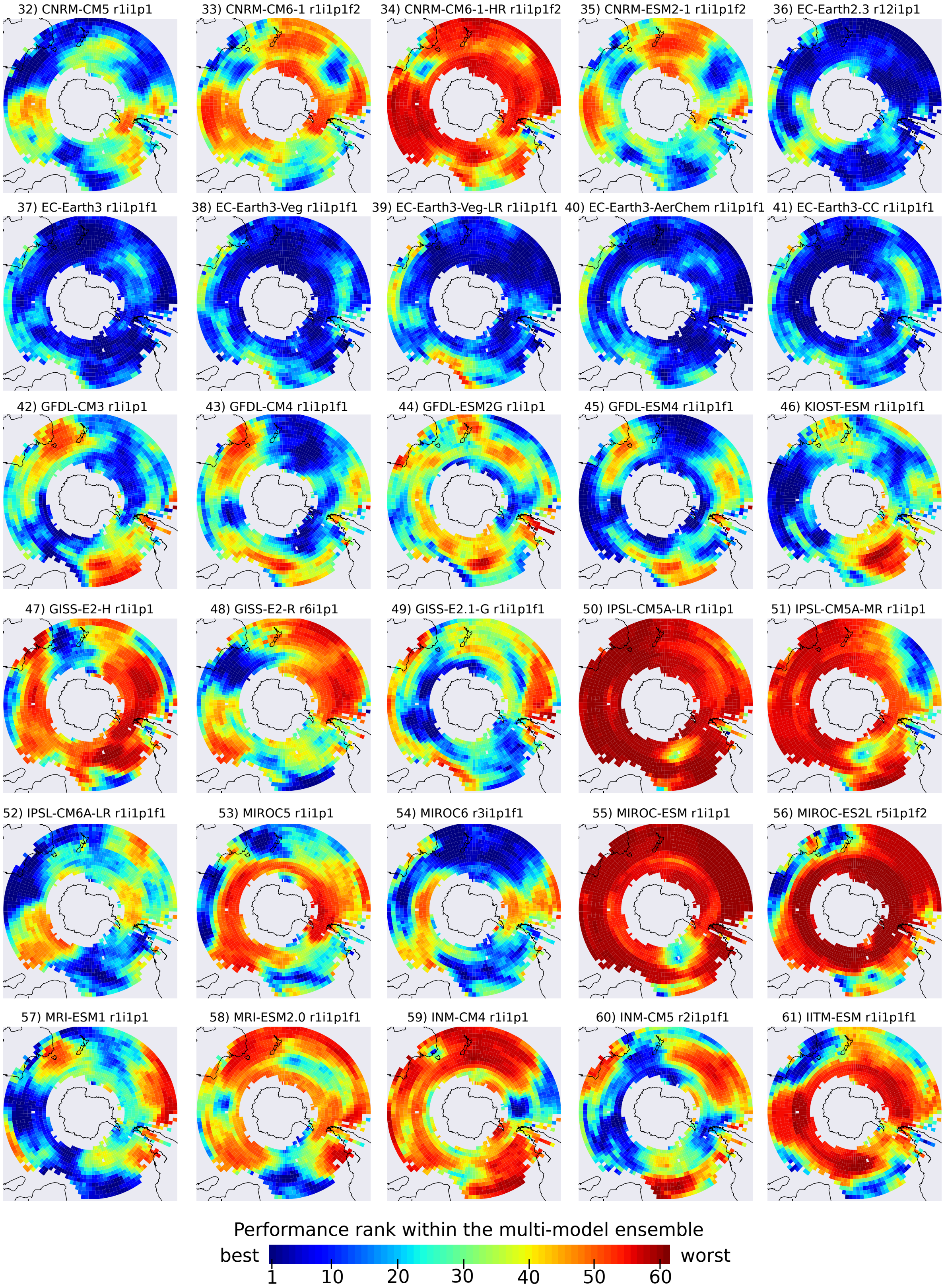

Figure 2. As Figure 1, part 2

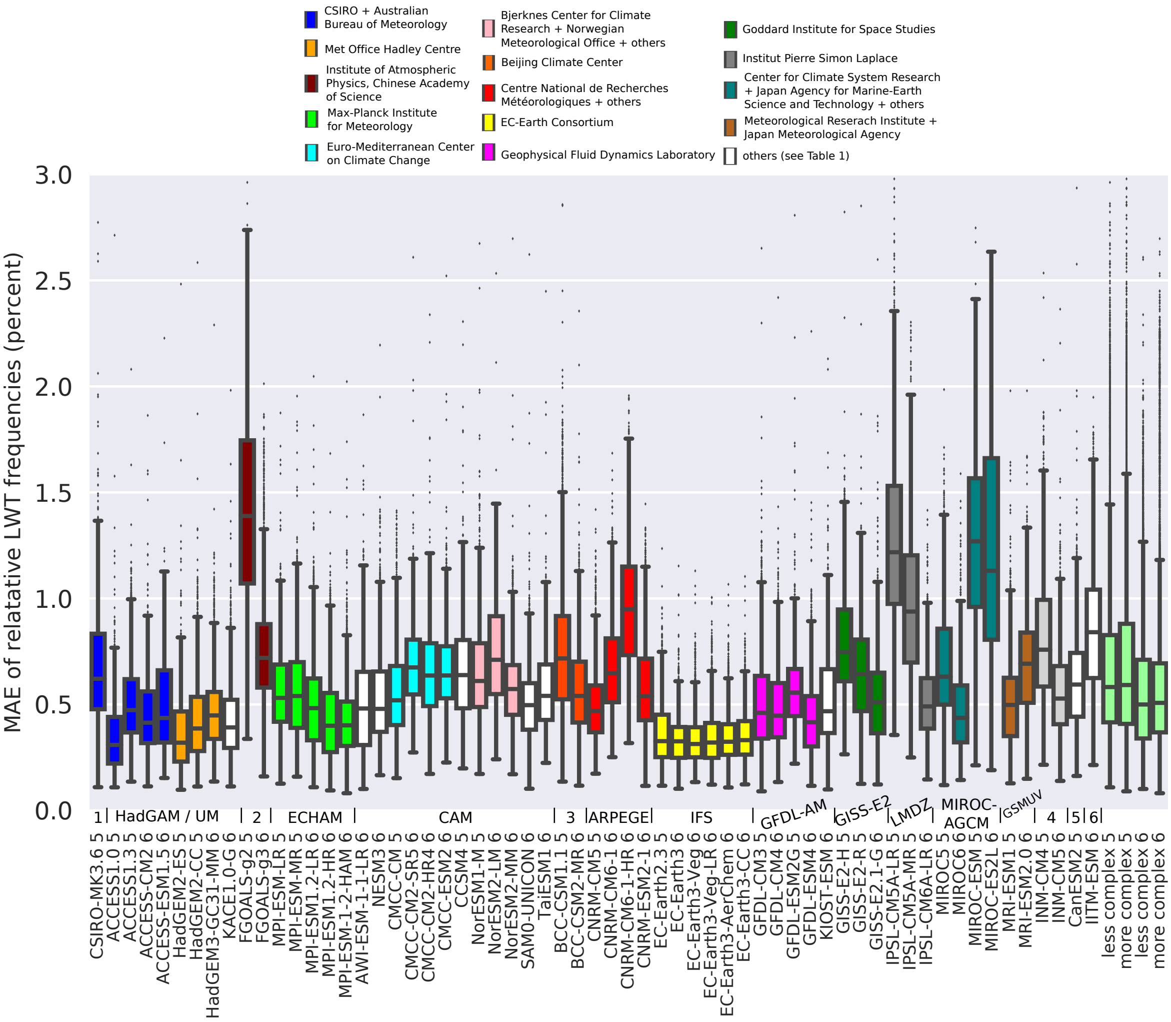

Figure 3. Summary model performance plot. Columns are constructed upon the model-specific, point-wise error values over the Southern Hemisphere as depicted in Figures 1 and 2. The four additional boxplots depicted in light green were built upon the joint error samples of the more and the less complex GCMs used in CMIP5 and 6, respectively. Colours refer to research institutes as listed in the legend. The acronyms of the coupled models, as well as their participation in either CMIP5 or 6 (indicated by the final integer) are shown below the X-axis. Above this axis, the atmospheric component of each coupled model is shown in addition. Results are for the 1979-2005 period and w.r.t. JRA-55. AGCM abbreviations along the X-axis are as defined as follows: 1) MK3-AGCM, 2) GAMIL, 3) BCC-AGCM, 4) INM-AM, 5) CanAM4 and 6) GFS; the names of the remaining AGCMs are indicated in the figure.

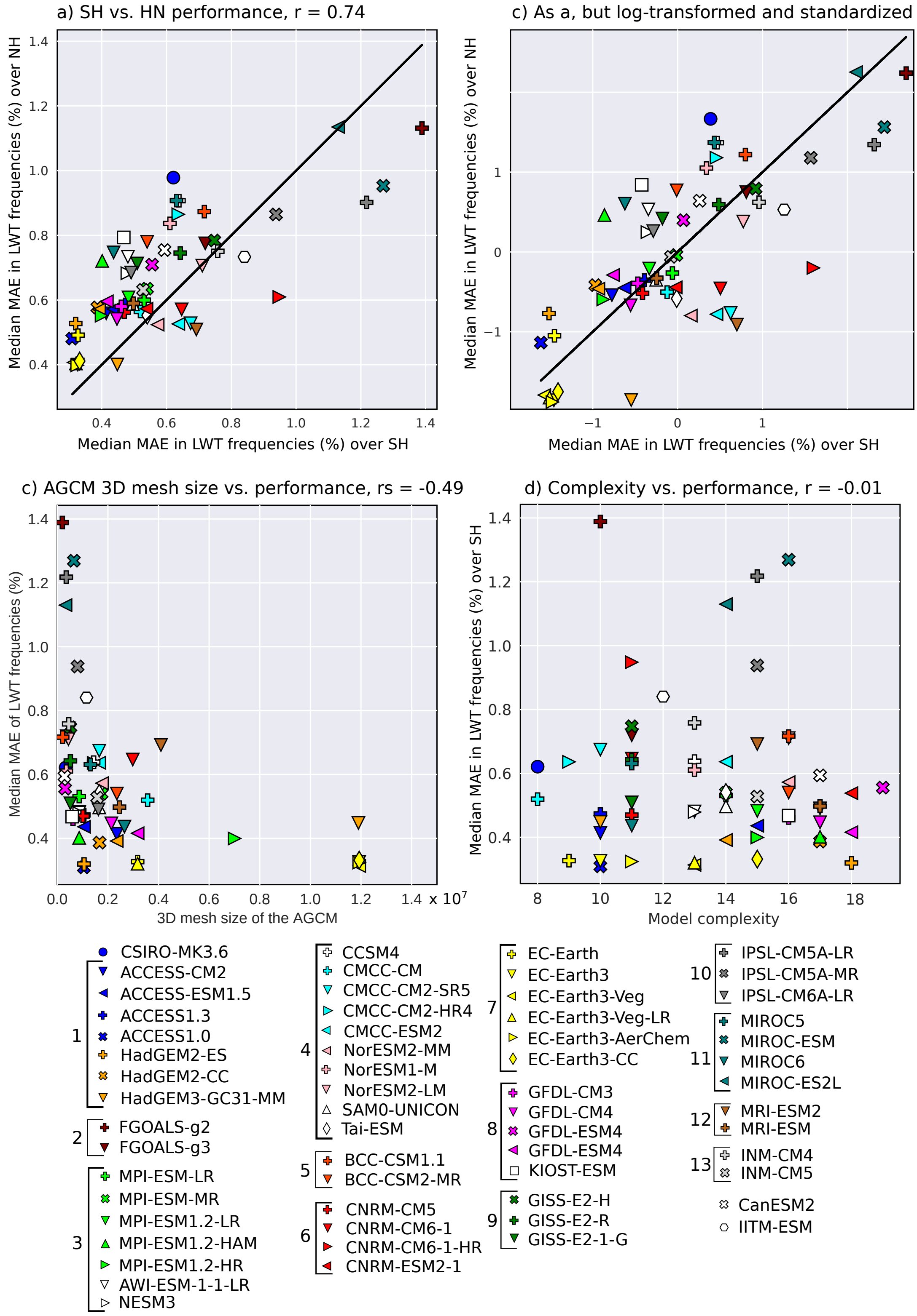

Figure 4. (a) Median model performance per GCM over the SH vs. NH; (b) As a, but for log-transformed and standardized data; (c) 3D mesh size of the AGCM vs. median model performance of the GCM in the SH and (d) Model complexity score proposed by Brands (2022a) vs. median model performance over the SH. AGCM families are indicated as follows: 1) HadGEM/UM, 2) GAMIL, 3) ECHAM, 4) CAM, 5) BCC-AGCM, 6) ARPEGE, 7) IFS, 8) GFDL-AM, 9) GISS-E2, 10) LMDZ, 11) MIROC/CCSR-AGCM, 12) GSMUV, 13) INM-AM

A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns

S. Brands[1], J.A. Fernández-Granja[2], J. Bedía[3,4], A. Casanueva[3,4], J. Fernández[2]

[1]MeteoGalicia, Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia, 15707 Santiago de Compostela, Spain
[2]Instituto de Física de Cantabria, Universidad de Cantabria-CSIC, 39005 Santander, Spain
[3]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, 39005 Santander, Spain
[4]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, 39005 Santander, Spain

Key Points:

- CMIP6 models perform better than CMIP5 models on average
- Southern Hemisphere model ranking similar to Northern Hemisphere ranking
- More complex model versions perform similar to less complex ones

Corresponding author: Swen Brands, `swen.brands@gmail.com`

Abstract

The performance of 61 global climate models participating in CMIP5 and 6 is evaluated for the Southern Hemisphere extratropics in terms of typical regional-scale atmospheric circulation patterns. These patterns are known to be linked with a number of key variables in atmospheric physics and chemistry and provide an overarching concept for model evaluation. First, hemispheric-wide error and ranking maps are provided for each model and regional details are described. Then, the results are compared with those obtained in a companion study for the Northern Hemisphere. For most models, the average error magnitude and ranking position is similar on both hemispheres, ruling out systematic tuning towards either of the two. CMIP6 models perform better on average than CMIP5 models and the interactive simulation of more climate system components does not deteriorate the results for most model families. Better performance is associated with higher resolution in the atmosphere, following a non-linear relationship.

Plain Language Summary

This letter provides a survey on the capability of global climate models to reproduce the regional atmospheric circulation in the Southern Hemisphere in present climate conditions. Climate models from the latest model generation are found to perform better on average than those of the previous generation and the obtained model ranking is similar to that found for the Northern Hemisphere in a companion study. While model performance is found to be generally unrelated to model complexity in terms of covered climate system components, better results are associated with higher model resolution in the atmosphere.

1 Introduction

The vast ocean and ice-sheet areas in the Southern Hemisphere (SH) extratropics are virtually inhabited, but play a key role for the global climate system and are thus of paramount importance for mankind. In this context, the quasi-persistent circumpolar westerly winds blowing along the open sea channel in the mid-latitudes are of key relevance for several reasons. Partly offset by meso-scale ocean eddies tending to break up the intense ocean stratification, the westerlies drive the up-welling of carbon and nutrient-rich deep water and also force the Antarctic Circumpolar Current (ACC), which is embedded in the Meridional Overturning Circulation that in turn governs the low frequency variability of the global climate system (Abernathey et al., 2011; Speer & Marshall, 2012; Meredith et al., 2012; Hogg et al., 2017; Zhang et al., 2019). Commonly described by the Southern Annular Mode, also referred to as "Antarctic Oscillation" (Trenberth, 1979; Rogers & van Loon, 1982; Thompson & Wallace, 2000), the westerlies have shifted poleward during the last decades while, simultaneously, the Hadley Cell and associated large-scale subsidence in the sub-tropics have intensified (Thompson et al., 2000; Nguyen et al., 2015; Fogt & Marshall, 2020). Both anomalies are projected to magnify during the course of the 21^{st} century in global climate model (GCM) experiments (Deng et al., 2022), leading to more frequent extreme events like, e.g., droughts (Holgate et al., 2020) or sea-surface warming events (Duran et al., 2020) whose accumulated effects also alter the mass balance of the glaciers and ice-sheets in the SH.

While Patagonian glaciers are mainly affected by temperature and precipitation anomalies (Boex et al., 2013), melting into the Amundsen and Bellingshausen Seas is the main driver of Antarctic continental ice loss (Hughes, 1981; Rignot et al., 2019). The aforementioned poleward shift of the westerlies has led to an enhanced transport of relatively warm Circumpolar Deep Water, located at intermediate depths below the cold surface ocean layer, towards the continental shelf of the aforementioned sea areas (Steig et al., 2012), thereby thinning the ice shelves from below and melting the glaciers and ice streams at their ground lines. Subject to large uncertainties (Rignot et al., 2011), these processes contribute to global sea-level rise (Fox-Kemper et al., 2021).

Over glacial-to-interglacial cycles, the strength and position of the westerlies are also considered key to variations in the upwelling of carbon-rich Antarctic Bottom Water (AABW) reservoirs, associated with CO_2 degassing into the atmosphere (Sigman & Boyle, 2000; Speer & Marshall, 2012). There are indications that strong westerlies located near the Antarctic continent —well aligned with the ACC—, typically occur during warm, interglacial periods and enhance the aforementioned process leading to an increase in global CO_2 concentrations. Weaker westerlies located far away from the Antarctic continent and poorly aligned with the ACC are, in turn, currently discussed to be characteristic of cold, glacial periods. CO_2 degassing into the atmosphere wold be reduced in this case, favouring a net carbon storage in the AABW (Toggweiler et al., 2006; Gray et al., 2021). Finally, AABW formation itself is also controlled by wind forcing, namely by southerly katabatic winds blowing down the Antarctic continent, pushing the sea-ice away from coast and thereby forming coastal polynyas. In these ocean water areas surrounded by sea-ice, the nutrient-rich upwelled waters are subject to brine rejection during sea-ice formation that leads to increase in salinity. Sinking to the ocean bottom is the consequence, where "preformed" nutrients can thereby accumulate. AABW formation is particularly productive in the Weddell and Ross Seas and is subject to pronounced low-frequency variability (Ito & Follows, 2005; Hogg et al., 2017; Silvano et al., 2020).

These considerations show that the atmospheric circulation in the SH extratropics, even in confined and relatively small regions such as the aforementioned sea areas, are relevant for the entire climate system. Consequently, comprehensive GCMs now extensively used in climate research should perform well in this regard.

The present study evaluates a large multi-model ensemble from the Coupled Model Intercomparison Projects 5 and 6 (Taylor et al., 2012; Eyring et al., 2016) in terms of the models' capability to reproduce the climatological frequencies of typical and recurrent patterns of the regional atmospheric circulation in the SH extratropics. To this aim, the Lamb Weather Types method (LWT), also known as Jenkison-Collison circulation typing approach (Lamb, 1972; Jenkinson & Collison, 1977; Jones et al., 1993) has been recently extended for systematic use in the SH (Fernández-Granja et al., 2023) and is here applied to 61 GCMs from CMIP5 and 6, and to 3 distinct reference reanalyses. The circulation types obtained from this method are known to correlate with many key variables in atmospheric physics and chemistry and therefore constitute an overarching concept to describe regional-scale climate variability. The method is thus complementary to those operating on larger scales used in a previous study (Bracegirdle et al., 2020) and it is a direct answer to the downscaling community's claim for process-based GCM evaluation based on the regional atmospheric circulation (Maraun et al., 2017; Røste & Landgren, 2022), here tailored to the SH mid-to-high latitudes (Olson et al., 2016; Fita et al., 2017; Charles et al., 2020; Evans et al., 2021). Together with the respective assessment for the Northern Hemisphere (Brands, 2022a), this study completes the picture of GCM performance in terms of regional atmospheric circulation patterns in the extratropics. Possible model tuning issues to either of the two hemispheres are also discussed.

2 Data and Methods

The study relies on 6-hourly instantaneous sea-level pressure data from 61 different GCM configurations participating in CMIP5 and 6, all retrieved from the ESGF data portals. *Historical* experiments are evaluated and the considered ensemble members for each GCM are listed in the *get_historical_metadata.py* function available from Brands et al. (2022). It will be shown that the role of internal model variability does not lead to substantial changes in the results (see Section 3.1). Since several EC-Earth model versions were found to be favoured when evaluated against ECMWF reanalyses (Dee et al., 2011; Hersbach et al., 2020) in the companion study conduced over the Northern Hemisphere (Brands, 2022a), the Japanese 55-year reanalysis (JRA-55) is here used as main reference dataset for model evaluation (Kobayashi et al., 2015).

The Jenkinson-Collison circulation types constitute an objective classification method based on the subjective approach made by Lamb (1972). This technique, also known as "Lamb Weather Types" (Jones et al., 1993), groups an instantaneous SLP pattern centered at a given grid-box into 27 classes depending on the direction of the geostrophic flow (or lack thereof) and the sign and strength of the vorticity. In addition to the purely cyclonic and anticyclonic types, there are 8 "purely directional" types —one for each of the 8 main cardinal directions— and 16 hybrid types characterized by a predominant flow from one of these directions combined with either cyclonic or anticyclonic conditions. A detailed description of the LWT method, including the extension to the SH relevant here, is provided in Fernández-Granja et al. (2023). The corresponding Python code is available from Brands et al. (2022). The LWT method is here applied in a rolling manner (Otero et al., 2017) looping through all boxes of a regular latitude-longitude grid covering a spatial domain extending from 30°S to 70°S with a 2.5° resolution. The considered time period is 1979 to 2005, for which data is available for all applied GCMs and reanalyses.

The LWT method is here said to be applicable for a given region if at least 20 types occur with a minimum relative frequency of 0.1% (i.e. n = 39 occurrences for 27 years and 6-hourly data) at the corresponding grid-box in the reference reanalysis (JRA-55). This criterion is fulfilled in virtually the entire study area. Moreover, to ensure that the regional-scale GCM ranking presented here is robust to a switch in the underlying reference reanalysis, ERA-Interim is evaluated against JRA-55 just as if it was another GCM and the obtained error is compared to the errors of the 61 authentic GCMs. If any of the considered GCMs is found to perform better than ERA-Interim at a given grid-box, this indicates large observational uncertainties in the corresponding region, leading to an exclusion of the grid-box from further assessment. Figure 1 shows that all grid-boxes over the Antarctic continent and adjacent sea-areas seasonally covered by sea-ice have to be excluded for this reason.

For consistency with Brands (2022a) and Brands (2022b), the mean absolute error (MAE) of the climatological relative frequencies of the $n = 27$ LWTs is used as main error measure at the grid-box scale:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |m_i - o_i| \tag{1}$$

where m_i and o_i are the relative frequencies of the i^{th} LWT (with $i = 1, 2, ..., 27$) from the GCM and reference reanalysis, respectively. Alternative error measures such as the *Transition Probability Matrix Score* (Fernandez-Granja et al., 2021) have been used as well, obtaining nearly identical results for the model ranking.

To explore the role of internal model variability due to initial conditions uncertainties, up to 18 distinct historical runs per GCM are evaluated for a subset of 13 GCMs, specified in Supplementary Figure 3. Then, the GCM complexity score proposed in Brands (2022a) is used, which is proportional to the number of Earth system components taken into account in the GCM and gives more weight on simulated than on prescribed components. This score is put into relation with the spatial median model performance over the SH in order to explore whether the more complex GCMs perform better or worse than the less complex ones on average. Finally, the SH results are plotted against the NH results obtained from Brands (2022a) in order to detect possible tuning efforts to either of the two hemispheres. For a comparison on equal terms, the NH results were modified by also removing the regions prone to substantial reanalysis uncertainties, as described above.

3 Results

3.1 Regional details

Figures 1 and 2 show the GCM ranking patterns based on the MAE for the 61 considered GCMs, with lower MAE values leading to better ranks and vice versa. The MAE values themselves are depicted in Supplementary Figures 1 and 2. Hereafter, individual GCMs will be grouped according to the atmosphere general circulation model (AGCM) used therein (Brands, 2022b).

From these figures, it can be seen that the IFS AGCM family (i.e. all EC-Earth GCM versions) performs best overall, followed by the HadGAM/UM AGCM family comprising the ACCESS and HadGEM GCMs, as well as KACE1.0G. All members of the HadGAM/UM family except ACCESS1.0 and HadGEM3-GC31-MM have similar error patterns, with larger errors in the Southern Ocean to the south, southwest and southeast of the Australian continent. ACCESS1.3, ACCESS-CM2, HadGEM3-GC31-MM and KACE1.0-G perform relatively poorly to the south and southwest of Cape of Good Hope and KACE1.0-G additionally performs badly off the east coast of South America. A large performance gain is observed from CSIRO-MK3.6 to the ACCESS GCM family, i.e. from the former to the present GCM family developed by CSIRO (see also Figure 3).

The GAMIL AGCM family comprising the FGOALS-g2 and g3 GCMs performs overly poorest in this multi-model comparison. The ECHAM AGCM family, including all MPI-ESM versions, AWI-ESM-1-1-LR, NESM3 and CMCC-CM, performs slightly worse than the IFS and HadGAM/UM families, except for MPI-ESM1-2-HR and MPI-ESM-1-2-HAM performing almost equally well. A particularly poor model performance is observed for NESM3 along virtually the entire subtropics, extending to the mid-latitudes in the South Atlantic Ocean, and for AWI-ESM-1-1-LR over the eastern South Pacific and eastern South Atlantic. The CAM AGCM family comprises the largest number of GCMs (CMCC-CM2-SR5, CMCC-CM2-HR4, CMCC-ESM2, CCSM4, NorESM1-M, NorESM2-LM, NorESM2-MM, SAM0-UNICON, TaiESM1, BCC-CSM1.1 and BCC-CSM2-MR) and yields intermediate to unfavourable ranks in most regions. CMCC-CM, NorESM2-MM and SAM0-UNICON perform best in this family, yielding very good ranks in specific regions. CanESM2 comprises the CanAM4 AGCM that is not used in any other GCM and performs relatively poorly.

For the ARPEGE AGCM family shown in Figure 2 (CNRM-CM5, CNRM-CM6-1, CNRM-CM6-1-HR and CNRM-ESM2-1), the model versions used in CMIP6 (CNRM-CM6-1, CNRM-CM6-1-HR and CNRM-ESM2-1) perform worse than the well performing CMIP5 version CNRM-CM5. Surprisingly, this decrease in model performance is particularly pronounced in the high resolution (HR) version. IFS (EC-Earth2.3, EC-Earth3, EC-Earth3-Veg, EC-Earth3-Veg-LR, EC-Earth3-AerChem and EC-Earth3-CC) is the best performing model family in the present study, obtaining very good ranks over a large fraction of the domain. Model ranks worse than 40 are very rare, except for the ocean area to the south of Africa in EC-Earth-Veg-LR. The performance of the GFDL-AM AGCM family comprising GFDL-CM3, GFDL-CM4, GFDL-ESM2G, GFDL-ESM4 and KIOST-ESM is similar in magnitude to the ECHAM family, with best results overall for GFDL-ESM4. In case of the GISS-E2 AGCM family, the use of the Russel ocean model in GISS-E2-R leads to substantially better results than the use of the HYCOM model used in GISS-E2-H, the configuration of these two GCMs being otherwise equal (Schmidt et al., 2014), and a further performance increase is obtained by the CMIP6 version GISS-E2.1-G, comparable to that obtained for the ECHAM and GFDL-AM families mentioned above (see Figure 3). The most pronounced performance gain from CMIP5 to 6 is obtained for the LMDZ AGCM family (i.e. from IPSL-CM5A-LR and MR to IPSL-CM6A-LR), yielding a MAE level for IPSL-CM6A-LR comparable to that obtained for the ECHAM and GFDL-AM families. The MIROC-AM family is prone to very large performance differences from the better performing versions MIROC5 and 6 to the substantially worse performing versions MIROC-ESM and MIROC-ES2L, which are both more complex (see Figure 4d). The performance of the GSMUV family decreases substantially from CMIP5 to 6 (from MRI-ESM1 and MRI-ESM2.0) whereas that of the INM-AM family increases drastically (from INM-CM4 to INM-CM5). Finally, IITM-ESM is one of the worst performing GCMs considered here, with large differences in the results from one region to another.

3.2 Performance Summary and Comparison with the NH results

In Figure 3, the hemispheric-wide MAE samples mapped in Supplementary Figures 1 and 2 are displayed in a single boxplot. Each item describes the error distribution of a specific GCM in terms of the median (horizontal black line), interquartile range (IQR, box) and whiskers extending to the full range, except for the outliers lying beyond $1.5 \times IQR$ below and above the 2^{nd} and 3^{rd} quartile, respectively. The last four boxes, depicted in light green, are built upon the joint samples of the more and the less complex GCMs used in CMIP5 and 6, respectively (outliers are not shown for these samples). To this end, the GCMs are grouped according to their complexity score obtained from Brands (2022a) and those obtaining a score ≥ 14 are considered more complex.

For both complexity classes, the models used in CMIP6 perform better on average than those used in CMIP5. The largest performance gains from CMIP5 to 6 are obtained for the FGOALS and IPSL GCMs. However, a performance loss is obtained for 4 GCM groups —ACCESS, CMCC-CM, CNRM-CM and MRI-ESM—, which are 2 groups more than for the NH results obtained in Brands (2022a). Supplementary Figure 3 shows that internal model variability does not substantially change the aforementioned results.

A comparison between the areal median performance in the SH vs. NH is provided in Figure 4a and b. Overall, GCM performance is better in the SH than in the NH (panel a), which may be simply due to the fact that GCMs tend to perform better over the ocean than over land (Brands, 2022a), the ocean area being much larger in the SH. A close correspondence is obtained for the median error samples of the two hemispheres, particularly if they are log-transformed and standardized separately in order to remove systematic differences in their hemisphere-specific shape, magnitude and dispersion (panel b). Largest deviations from the diagonal are obtained for CNRM-CM6-1-HR, MRI-ESM2, CMCC-CM2-SR5, CMCC-ESM2 and HadGEM3-GC31-MM, performing better over the NH, and for CSIRO-MK3.6, KIOST-ESM, GISS-E2-1-G, MPI-ESM1.2-HAM and INM-CM5, performing better over the SH.

A significant non-linear relationship is obtained between the resolution of the AGCM —here described by the number of grid-points constituting the global 3-dimensional mesh (longitudes × latitudes × vertical layers)—, and the median model performance, obtaining a Spearman correlation coefficient (rs) of -0.49. Higher resolution is associated with better performance, particularly above a threshold of approximately 1.8×10^7 grid-boxes. Note that CNRM-CM6-1-HR and CNRM-ESM2-1 are not shown in Figure 4c because they are out-of-scale due to their very high resolution. Interestingly, the corresponding link with the 3D resolution of the ocean sub-model is weak (rs = -0.29), yet significant at a test level of 5%.

Finally, median model performance over the SH is generally not associated with model complexity (r = -0.01) and, for most model families, the more complex versions perform at least equally well than the less complex ones (see Figure 4d). The MIROC-AGCM family is an exception in this sense, since the more complex model versions MIROC-ESM and MIROC-ES2L, probably due to their low horizontal resolution in the atmosphere (T42) (Brands et al., 2022), perform substantially worse than the less complex versions MIROC5 and 6 (T85).

4 Conclusions

In the present study, 61 different GCMs from CMIP5 and 6 have been evaluated in the SH extratropics excluding Antarctica, focusing on the models' ability to reproduce the climatological frequency of the 27 Lamb Weather Types, known to be associated with many environmental variables and thus constituting a overarching concept to regional-scale climate variability.

While all of the model families performing poorly in CMIP5 have improved considerably in CMIP6, most of the families already performing well in CMIP5 have suffered a slight performance loss. For most model families, the spatial average performance for the SH is similar to that obtained for the NH (Brands, 2022a), suggesting that systematic model tuning to either of the two hemispheres can be ruled out in general terms. For a small number of specific GCMs, however, substantial performance differences are obtained from one hemisphere to another and the reasons for this should be assessed in future studies. Whereas a higher resolution in the atmospheric sub-model of the considered GCMs is found to be associated with better performance, following an exponentially decreasing relationship, GCM complexity as defined in Brands (2022a) is generally unrelated to performance, except for the MIROC-AGCM family, whose more complex versions perform worse than the less complex ones over the SH. This is a promising result since the more complex models are also prone to more error sources. It is also an argument for the use of the more complex models, as they provide a more complete picture of the feedback processes governing the climate system (Séférian et al., 2019; Dunne et al., 2020; Döscher et al., 2021).

Open Research Section

Supplementary Figures 1 to 3 are contained in the Supporting Information (SI) file to this article, available from GRL's homepage. The Python source code underlying this study and the GCM metadata archive *get_historitcal_metadata.py* are publicly available from Brands et al. (2022) and so is the LWT dataset for the considered GCMs and reanalyses, retrievable from Brands et al. (2023b). Additional auxiliary material containing 1) separate pdf files for each error and ranking map, 2) netCDF files containing grid-box-scale GCM errors and 3) summary csv files listing the model complexity score from Brands (2022a) as well as the spatial median performance over the SH domain for each GCM can be retrieved from Brands et al. (2023a).

Acknowledgments
The authors acknowledge the public availability of the CMIP datasets via the ESGF data portals, as well the free distribution of the ECMWF and JMA reanalysis products. This study is part of the I+D+i project CORDyS (PID2020-116595RB-I00), funded by MCIN/AEI/10.13039/501100011033. J.A.F. acknowledges support from project ATLAS (PID2019-111481RB-I00), grant PRE2020-094728 funded by MCIN/AEI/10.13039/501100011033 and ESF investing in your future. A.C. acknowledges support from Project COMPOUND (TED2021-131334A-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. J.B. thanks his parents for their parental care in the early years of his life. S.B. would like to thank CESGA and AMTEGA for providing computational resources.

References

Abernathey, R., Marshall, J., & Ferreira, D. (2011). The dependence of Southern Ocean meridional overturning on wind stress. *Journal of Physical Oceanography*, *41*(12), 2261 - 2278. doi: 10.1175/JPO-D-11-023.1

Boex, J., Fogwill, C., Harrison, S., Glasser, N., Hein, A., Schnabel, C., & Xu, S. (2013). Rapid thinning of the Late Pleistocene Patagonian Ice Sheet followed migration of the Southern Westerlies. *Scientific reports*, *3*, 2118. doi: 10.1038/srep02118

Bracegirdle, T. J., Holmes, C. R., Hosking, J. S., Marshall, G. J., Osman, M., Patterson, M., & Rackow, T. (2020). Improvements in circumpolar Southern Hemisphere extratropical atmospheric circulation in CMIP6 compared to CMIP5. *Earth and Space Science*, *7*(6), e2019EA001065. doi: https://doi.org/10.1029/2019EA001065

Brands, S. (2022a). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. doi: https://doi.org/10.5194/gmd-15-1375-2022

Brands, S. (2022b). Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble. *Geophysical Research Letters*, *49*(23), e2022GL101446. doi: https://doi.org/10.1029/2022GL101446

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023a). Auxiliary online material to Brands et al. (2023): A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns. *figshare*. doi: 10.6084/m9.figshare.22193443.v1

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023b). Southern Hemisphere Lamb Weather Types from historical GCM experiments and various reanalyses (1.0) [data set]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.7612988

Brands, S., Tatebe, H., Danek, C., Fernández, J., Swart, N., Volodin, E., ... Tongwen, W. (2022). Python code to calculate Lamb circulation types for the Northern Hemisphere derived from historical CMIP simulations and reanalysis data [code]. *Zenodo*. doi: https://doi.org/10.5281/zenodo.4555367

Charles, S. P., Chiew, F. H. S., Potter, N. J., Zheng, H., Fu, G., & Zhang, L. (2020). Impact of downscaled rainfall biases on projected runoff changes. *Hydrology and Earth System Sciences*, *24*(6), 2981–2997. doi: 10.5194/hess-24-2981-2020

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., … Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, *137*(656, Part a), 553-597. doi: 10.1002/qj.828

Deng, K., Azorin-Molina, C., Yang, S., Hu, C., Zhang, G., Minola, L., & Chen, D. (2022). Changes of Southern Hemisphere westerlies in the future warming climate. *Atmospheric Research*, *270*, 106040. doi: 10.1016/j.atmosres.2022.106040

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., … Zhang, Q. (2021). The EC-Earth3 Earth System Model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development Discussions*, *2021*, 1–90. doi: 10.5194/gmd-2020-446

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., … Zhao, M. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. doi: https://doi.org/10.1029/2019MS002015

Duran, E. R., England, M. H., & Spence, P. (2020). Surface ocean warming around Australia driven by interannual variability and long-term trends in Southern Hemisphere westerlies. *Geophysical Research Letters*, *47*(9), e2019GL086605. doi: https://doi.org/10.1029/2019GL086605

Evans, J., Virgilio, G., Hirsch, A., Hoffmann, P., Remedio, A. R., Ji, F., … Coppola, E. (2021). The CORDEX-Australasia ensemble: evaluation and future projections. *Climate Dynamics*, *57*, 1385–1401. doi: 10.1007/s00382-020-05459-0

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: https://doi.org/10.5194/gmd-9-1937-2016

Fernandez-Granja, J. A., Casanueva, A., Bedia, J., & Fernández, J. (2021). Improved atmospheric circulation over Europe by the new generation of CMIP6 Earth System Models. *Climate Dynamics*, *56*, 3527–3540. doi: https://doi.org/10.1007/s00382-021-05652-9

Fernández-Granja, J. A., Brands, S., Bedia, J., Casanueva, A., & Fernández, J. (2023). Exploring the limits of the Jenkinson–Collison weather types classification scheme: a global assessment based on various reanalyses. *Climate Dynamics*. doi: 10.1007/s00382-022-06658-7

Fita, L., Evans, J., Argueso, D., King, A., & Liu, Y. (2017). Evaluation of the regional climate response in Australia to large-scale climate modes in the historical NARCliM simulations. *Climate Dynamics*, *49*, 1-15. doi: 10.1007/s00382-016-3484-x

Fogt, R. L., & Marshall, G. J. (2020). The Southern Annular Mode: Variability, trends, and climate impacts across the Southern Hemisphere. *WIREs Climate Change*, *11*(4), e652. doi: https://doi.org/10.1002/wcc.652

Fox-Kemper, B., Hewitt, H., Xiao, C., Aoalgeirsdottir, G., Drijfhout, S. S., Edwards, T., … Yu, Y. (2021). Climate Change 2021: The Physical Science Basis. contribution of Working Group i to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In V. Masson-Delmotte et al. (Eds.), (pp. 1211–1362). Cambridge University Press. doi: 10.1017/9781009157896.011

Gray, W., de Lavergne, C., Jnglin Wills, R., Menviel, L., Spence, P., Holzer, M., … Michel, E. (2021). Poleward shift in the Southern Hemisphere westerly winds synchronous with the deglacial rise in CO2. *Research Square*, 1–50. doi: 10.21203/rs.3.rs-404786/v1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hogg, A. M., Spence, P., Saenko, O. A., & Downes, S. M. (2017). The energetics of Southern Ocean upwelling. *Journal of Physical Oceanography*, *47*(1), 135 - 153. doi: 10.1175/JPO-D-16-0176.1

Holgate, C. M., Van Dijk, A. I. J. M., Evans, J. P., & Pitman, A. J. (2020). Local and remote drivers of southeast Australian drought. *Geophysical Research Letters*, *47*(18), e2020GL090238. doi: https://doi.org/10.1029/2020GL090238

Hughes, T. J. (1981). The weak underbelly of the West Antarctic ice sheet. *Journal of Glaciology*, *27*(97), 518–525. doi: 10.3189/S002214300001159X

Ito, T., & Follows, M. (2005). Preformed phosphate, soft tissue pump and atmospheric CO2. *Journal of Marine Research*, *63*, 813–839. doi: 10.1357/0022240054663231

Jenkinson, A., & Collison, F. (1977). *An initial climatology of gales over the North Sea* (Vol. 62; Tech. Rep. No. 18). Bracknell, UK: Meterololgical Office.

Jones, P. D., Hulme, M., & Briffa, K. R. (1993). A comparison of Lamb circulation types with an objective classification scheme. *International Journal of Climatology*, *13*(6), 655-663. doi: https://doi.org/10.1002/joc.3370130606

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 Reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5-48. doi: 10.2151/jmsj.2015-001

Lamb, H. (1972). British Isles weather types and a register of daily sequence of circulation patterns, 1861-1971. *Geophysical Memoir*, *116*, 85pp. (HMSO)

Maraun, D., Shepherd, T., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J., … Mearns, L. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*, 764-773. doi: https://doi.org/10.1038/nclimate3418

Meredith, M. P., Garabato, A. C. N., Hogg, A. M., & Farneti, R. (2012). Sensitivity of the overturning circulation in the Southern Ocean to decadal changes in wind forcing. *Journal of Climate*, *25*(1), 99 – 110. doi: 10.1175/2011JCLI4204.1

Nguyen, H., Lucas, C., Evans, A., Timbal, B., & Hanson, L. (2015). Expansion of the Southern Hemisphere Hadley Cell in response to greenhouse gas forcing. *Journal of Climate*, *28*(20), 8067 - 8077. doi: 10.1175/JCLI-D-15-0139.1

Olson, R., Evans, J., Di Luca, A., & Argueso, D. (2016). The NARCliM project: Model agreement and significance of climate projections. *Climate Research*, *69*, 209-227. doi: 10.3354/cr01403

Otero, N., Sillmann, J., & Butler, T. (2017). Assessment of an extended version of the Jenkinson-Collison classification on CMIP5 models over europe. *Climate Dynamics*, 1559–1579. doi: 10.1007/s00382-017-3705-y

Rignot, E., Mouginot, J., Scheuchl, B., van den Broeke, M., van Wessem, M. J., & Morlighem, M. (2019). Four decades of Antarctic Ice Sheet mass balance from 1979-2017. *Proceedings of the National Academy of Sciences*, *116*(4), 1095-1103. doi: 10.1073/pnas.1812883116

Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A., & Lenaerts, J. T. M. (2011). Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophysical Research Letters*, *38*(5), L05503. doi: 10.1029/2011GL046583

Rogers, J. C., & van Loon, H. (1982). Spatial variability of sea level pressure and 500 mb height anomalies over the Southern Hemisphere. *Monthly Weather Review*, *110*(10), 1375 - 1392. doi: 10.1175/1520-0493(1982)110⟨1375:SVOSLP⟩2.0.CO;2

Røste, J., & Landgren, O. (2022). Impacts of dynamical downscaling on circulation type statistics in the Euro-CORDEX ensemble. *Climate Dynamics*, *59*, 2445–2466. doi: 10.1007/s00382-022-06219-y

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., … Zhang, J. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, *6*(1), 141-184. doi: 10.1002/2013MS000265

Sigman, D., & Boyle, E. (2000). Glacial/interglacial variations in atmospheric carbon dioxide. *Nature*, *407*, 859-69. doi: 10.1038/35038000

Silvano, A., Foppert, A., Rintoul, S., Holland, P., Tamura, T., Kimura, N., … Macdonald, A. (2020). Recent recovery of Antarctic Bottom Water formation in the Ross Sea driven by climate anomalies. *Nature Geoscience*, *13*, 1-7. doi: 10.1038/s41561-020-00655-3

Speer, K., & Marshall, J. (2012). Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geoscience - NAT GEOSCI*, *5*, 171-180. doi: 10.1038/ngeo1391

Steig, E., Ding, Q., Battisti, D., & Jenkins, A. (2012). Tropical forcing of Circumpolar Deep Water Inflow and outlet glacier thinning in the Amundsen Sea Embayment, West Antarctica. *Annals of Glaciology*, *53*(60), 19–28. doi: 10.3189/2012AoG60A110

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., … Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182-4227. doi: 10.1029/2019MS001791

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485-498. doi: https://doi.org/10.1175/BAMS-D-11-00094.1

Thompson, D. W. J., & Wallace, J. M. (2000). Annular modes in the extratropical circulation. part I: Month-to-month variability. *Journal of Climate*, *13*(5), 1000 - 1016. doi: 10.1175/1520-0442(2000)013⟨1000:AMITEC⟩2.0.CO;2

Thompson, D. W. J., Wallace, J. M., & Hegerl, G. C. (2000). Annular modes in the extratropical circulation. part II: Trends. *Journal of Climate*, *13*(5), 1018 - 1036. doi: 10.1175/1520-0442(2000)013⟨1018:AMITEC⟩2.0.CO;2

Toggweiler, J. R., Russell, J. L., & Carson, S. R. (2006). Midlatitude westerlies, atmospheric CO2, and climate change during the ice ages. *Paleoceanography*, *21*(2). doi: https://doi.org/10.1029/2005PA001154

Trenberth, K. E. (1979). Interannual variability of the 500 mb zonal mean flow in the Southern Hemisphere. *Monthly Weather Review*, *107*(11), 1515 - 1524. doi: 10.1175/1520-0493(1979)107⟨1515:IVOTMZ⟩2.0.CO;2

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., … Little, C. M. (2019). A review of the role of the Atlantic Meridional Overturning Circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*(2), 316-375. doi: 10.1029/2019RG000644

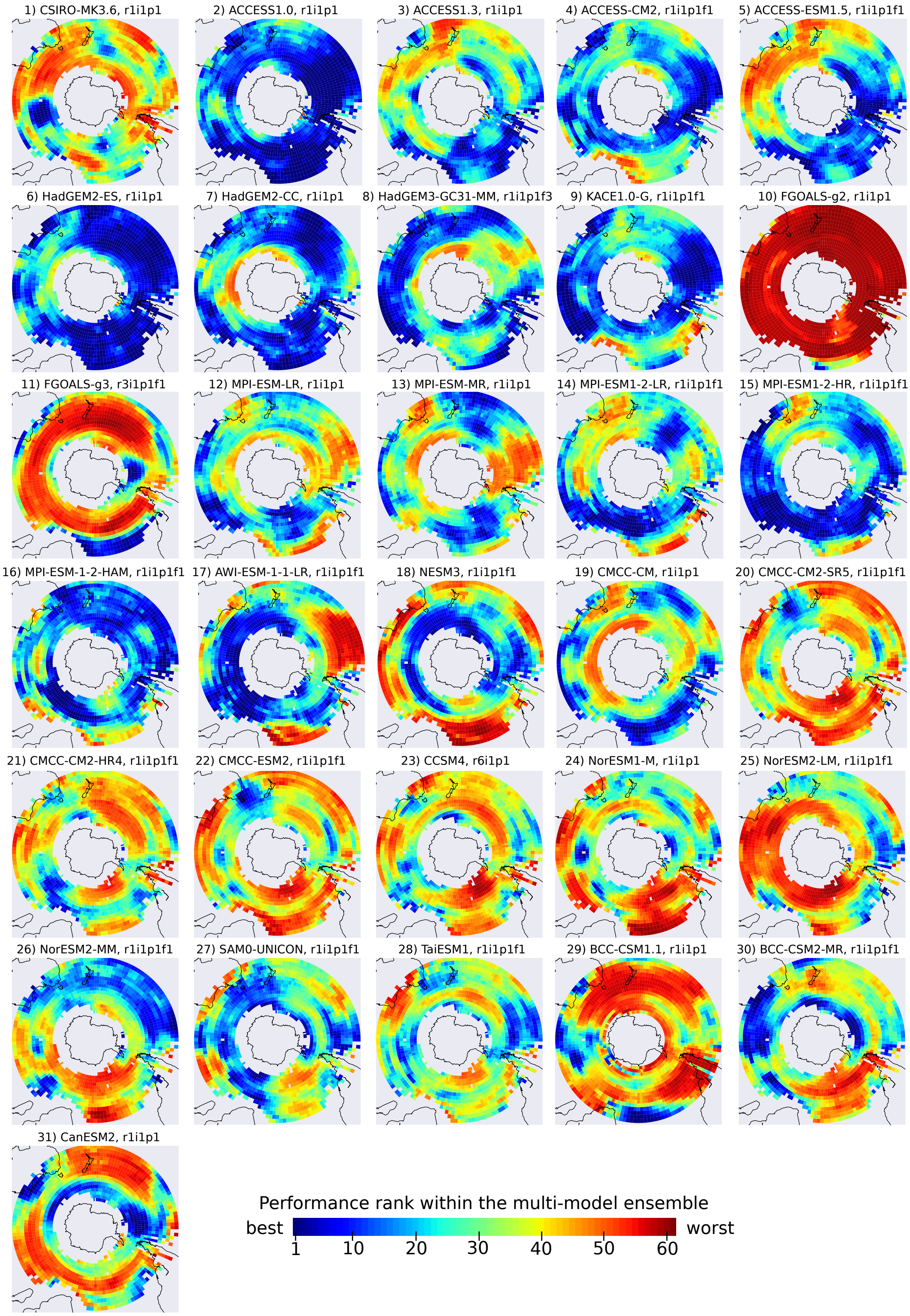

Figure 1. Ranking of the GCMs according to MAE defined in Equation 1, reference: JRA-55, 1979-2005, part 1

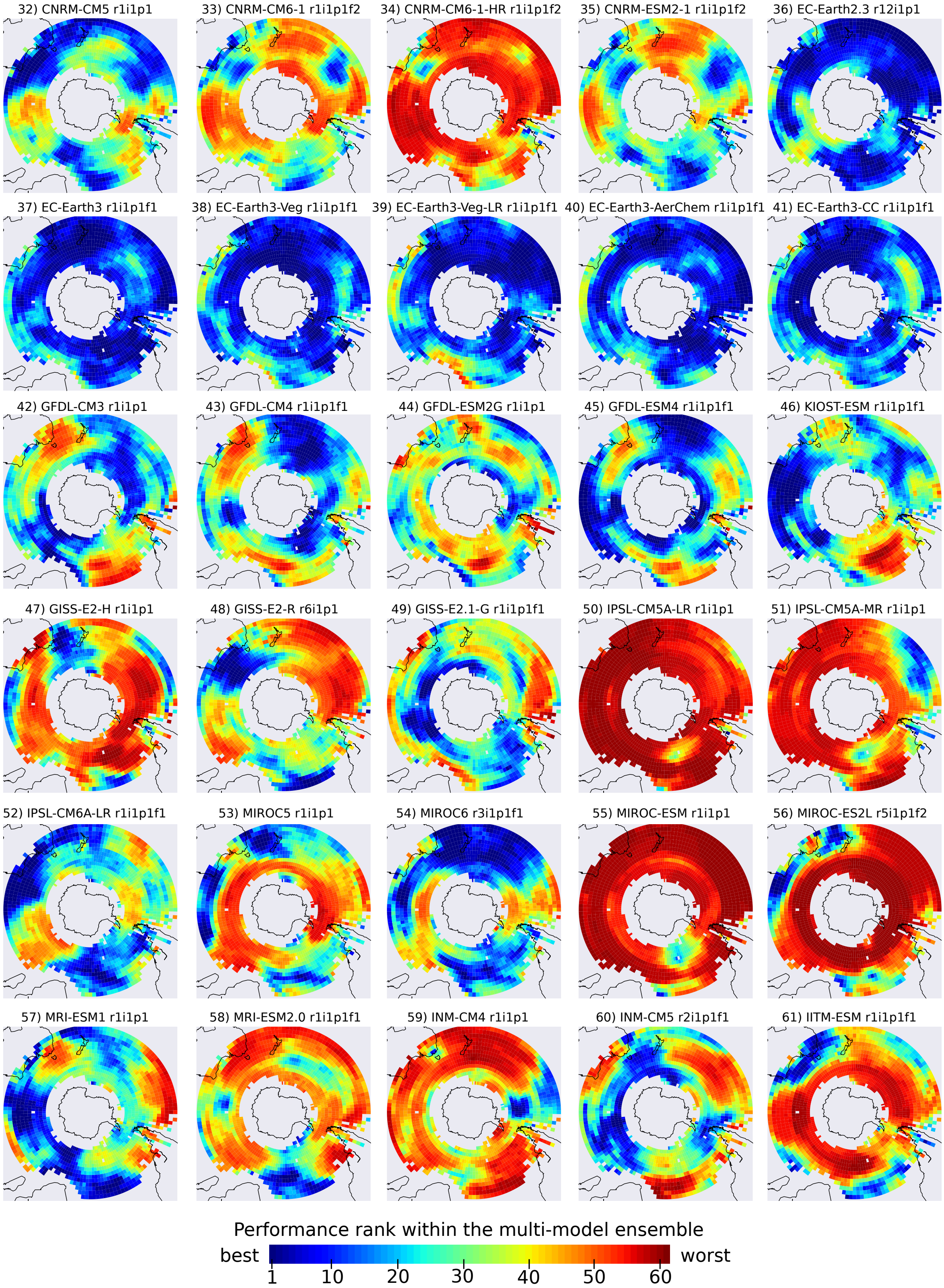

Figure 2. As Figure 1, part 2

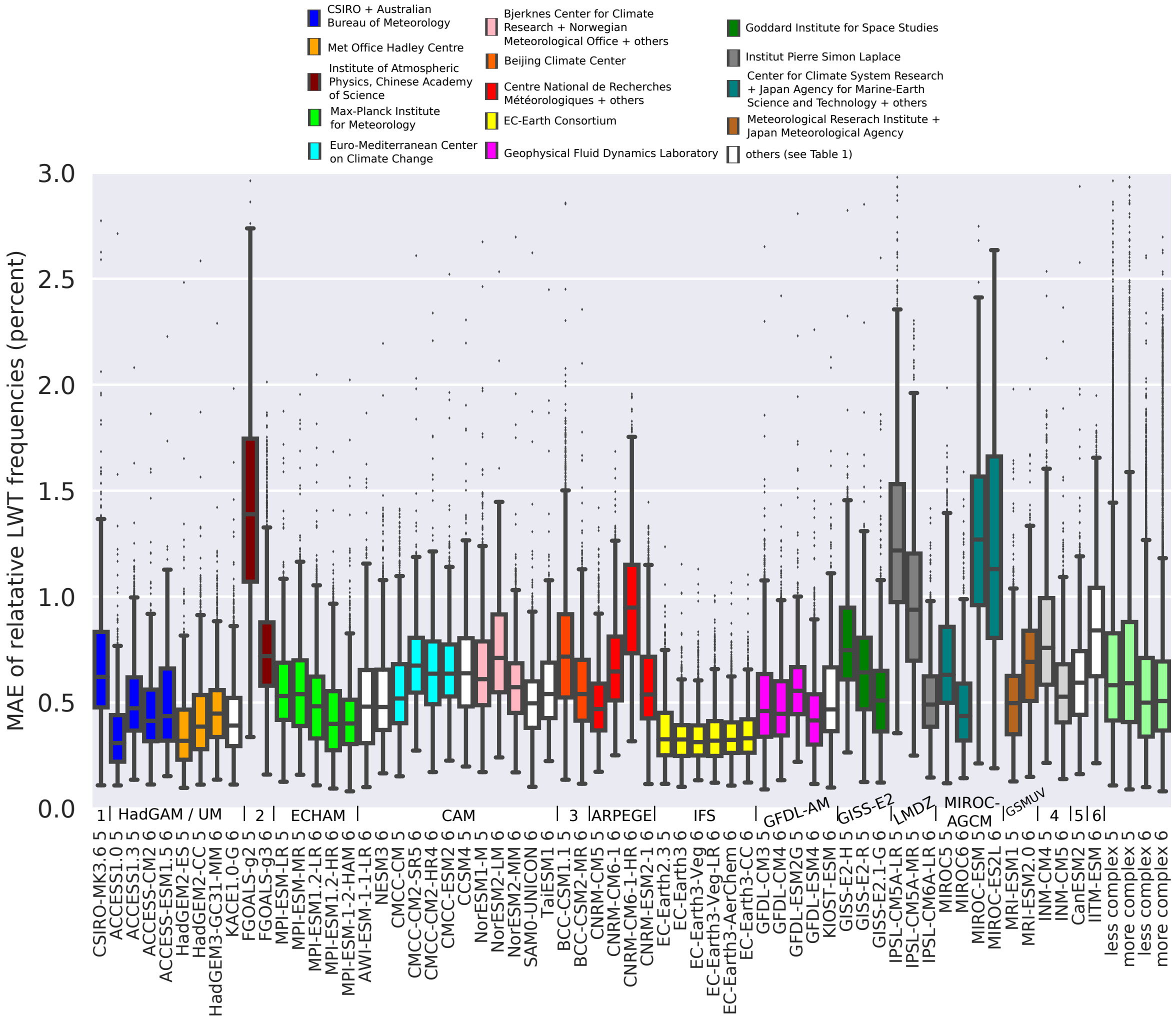

Figure 3. Summary model performance plot. Columns are constructed upon the model-specific, point-wise error values over the Southern Hemisphere as depicted in Figures 1 and 2. The four additional boxplots depicted in light green were built upon the joint error samples of the more and the less complex GCMs used in CMIP5 and 6, respectively. Colours refer to research institutes as listed in the legend. The acronyms of the coupled models, as well as their participation in either CMIP5 or 6 (indicated by the final integer) are shown below the X-axis. Above this axis, the atmospheric component of each coupled model is shown in addition. Results are for the 1979-2005 period and w.r.t. JRA-55. AGCM abbreviations along the X-axis are as defined as follows: 1) MK3-AGCM, 2) GAMIL, 3) BCC-AGCM, 4) INM-AM, 5) CanAM4 and 6) GFS; the names of the remaining AGCMs are indicated in the figure.

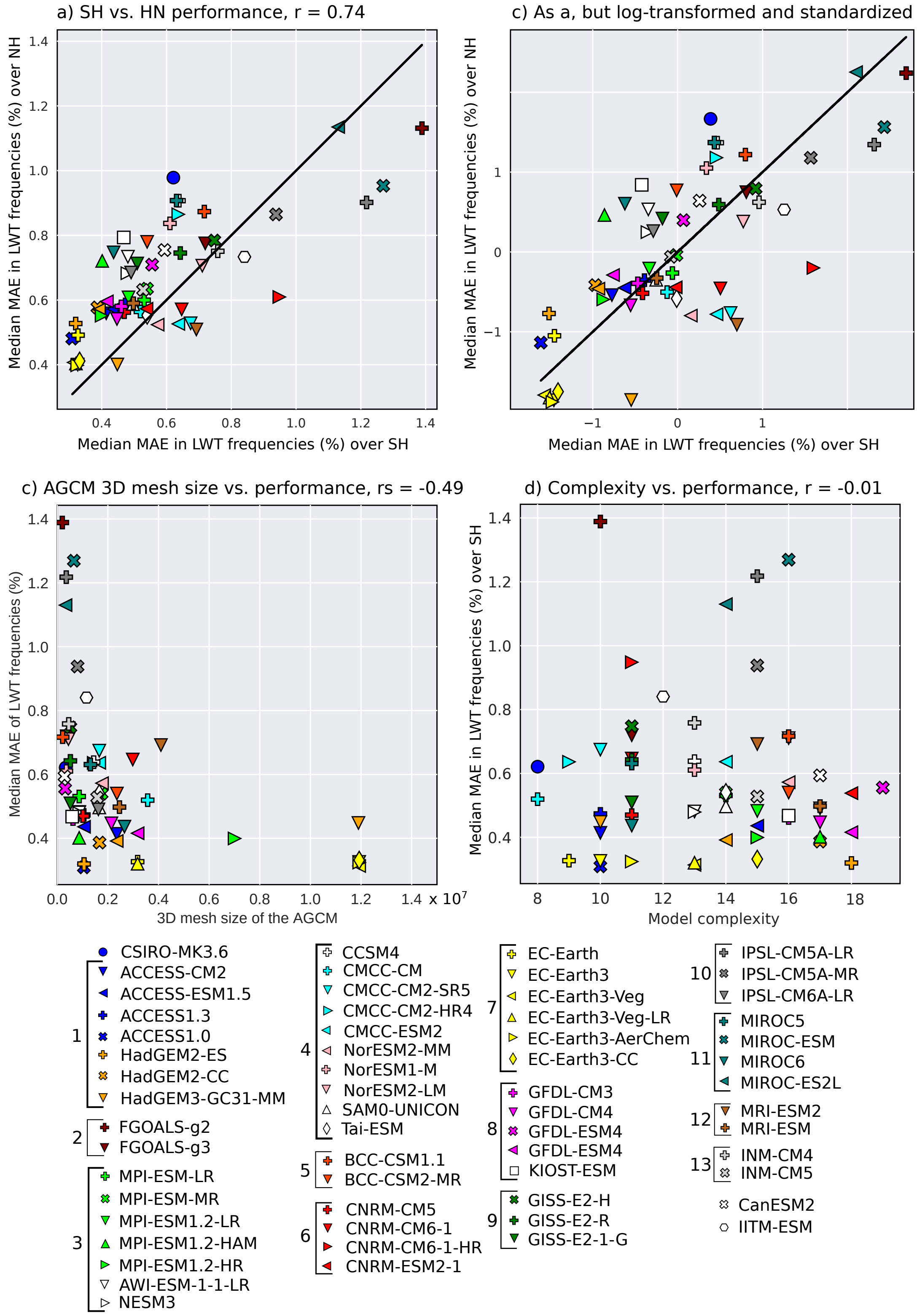

Figure 4. (a) Median model performance per GCM over the SH vs. NH; (b) As a, but for log-transformed and standardized data; (c) 3D mesh size of the AGCM vs. median model performance of the GCM in the SH and (d) Model complexity score proposed by Brands (2022a) vs. median model performance over the SH. AGCM families are indicated as follows: 1) HadGEM/UM, 2) GAMIL, 3) ECHAM, 4) CAM, 5) BCC-AGCM, 6) ARPEGE, 7) IFS, 8) GFDL-AM, 9) GISS-E2, 10) LMDZ, 11) MIROC/CCSR-AGCM, 12) GSMUV, 13) INM-AM

Supporting Information for ”A global climate model performance atlas for the Southern Hemisphere extratropics based on regional atmospheric circulation patterns”

S. Brands[1], J.A. Fernández-Granja[2], J. Bedía[3,4], A. Casanueva[3,4], J. Fernández[2]

[1]MeteoGalicia, Consellería de Medio Ambiente, Territorio y Vivienda - Xunta de Galicia, 15707 Santiago de Compostela, Spain

[2]Instituto de Física de Cantabria, Universidad de Cantabria-CSIC, 39005 Santander, Spain

[3]Dept. Matemática Aplicada y Ciencias de la Computación (MACC), Universidad de Cantabria, 39005 Santander, Spain

[4]Grupo de Meteorología y Computación, Universidad de Cantabria, Unidad Asociada al CSIC, 39005 Santander, Spain

Content of the present PDF file:

- Supplementary Figures 1 to 3

In addition to these figures, an **auxiliary data repository** to this work has been created at `https://doi.org/10.6084/m9.figshare.22193443.v1`

, see Brands et al. (2023a) in the main article file. A detailed description of the repository can be found in the the **README.txt** file included therein. Among other things, this repository also contains separate pdf files for each map provided in this study. Contact: swen.brands@gmail.com

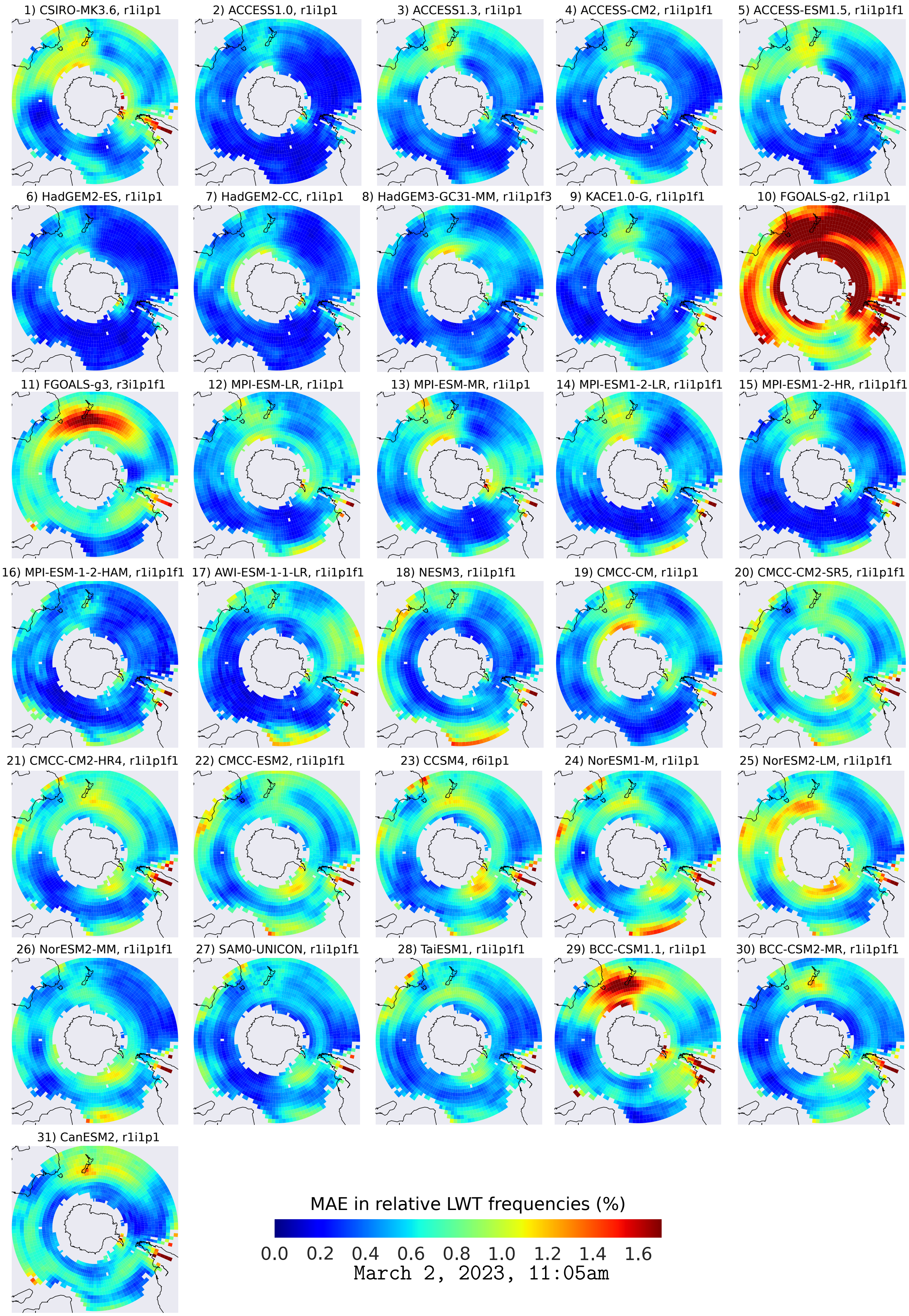

Figure S1. As Figure 1 in the article file, but showing the MAE instead.

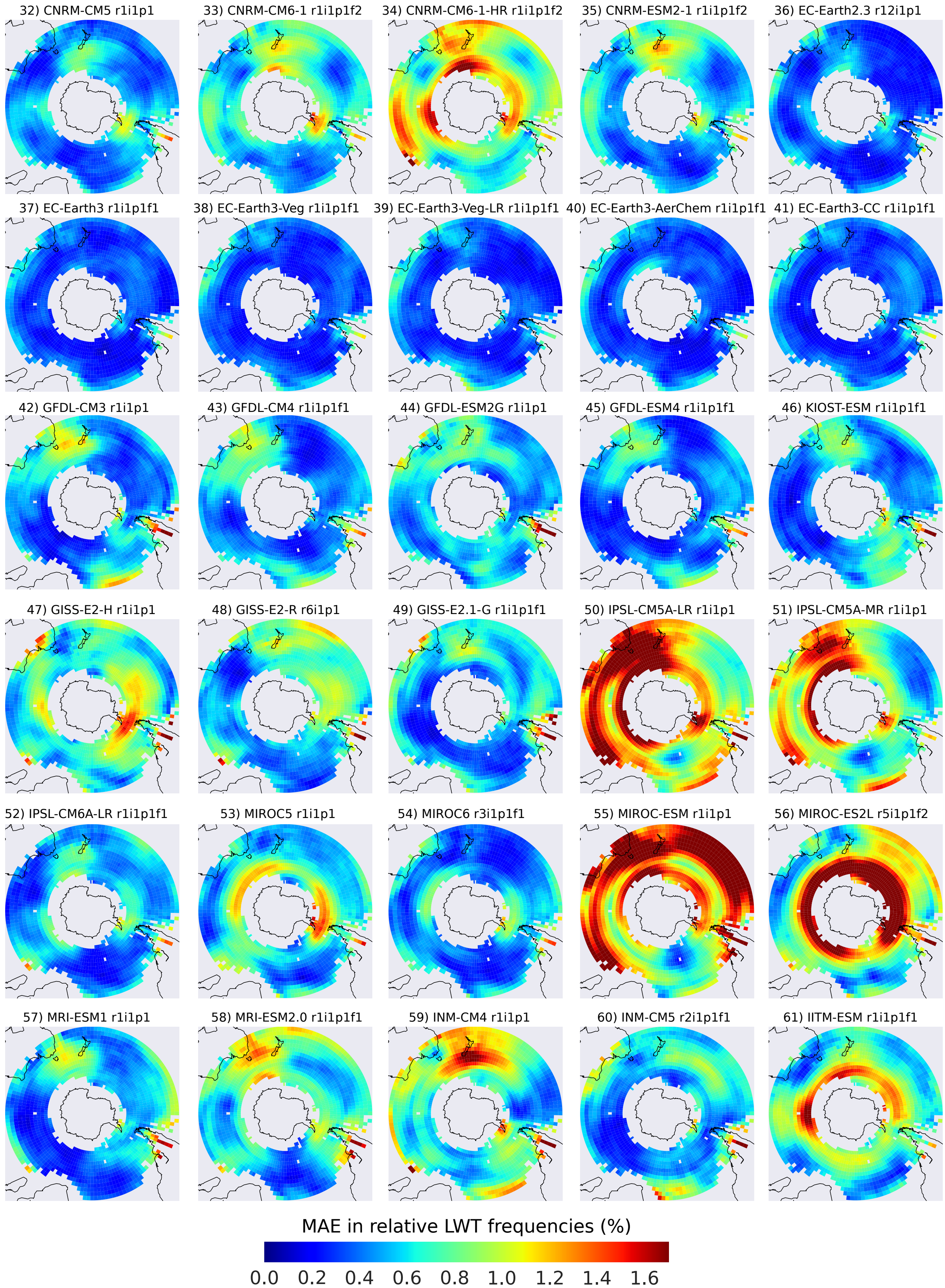

Figure S2. As Figure 2 in the article file, but showing the MAE instead.

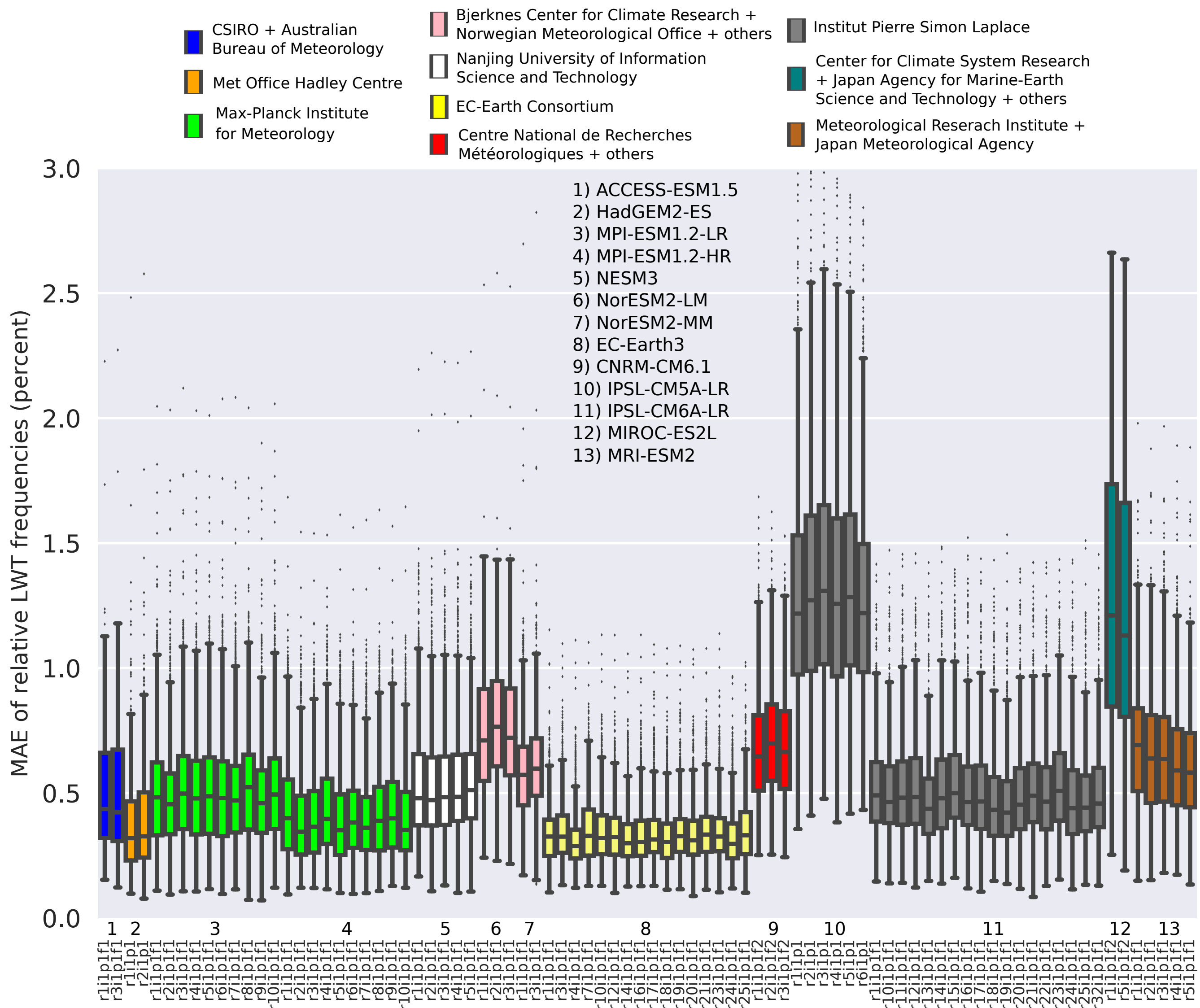

Figure S3. As Figure 3 in the article file, but considering 70 additional runs for a subset of 13 distinct coupled models. The colours referring to the coordinating research institute are identical to Figure 3, except for the *Nanjing University of Information Science and Technology* painted white. Up to 2 ensembles per institute are shown and the acronyms of the individual coupled models are indicated by numbers. The exact run specifications are provided along the x-axis.