CMIP6 model fidelity in capturing the Southern Hemisphere storm track and its connections with low-frequency variability

Isaac Campbell[1] and James Arthur Renwick[2]

[1]Te Herenga Waka - Victoria, University of Wellington
[2]Victoria University of Wellington

March 13, 2023

Abstract

Storm tracks are a key component of global atmospheric circulation. Their influence ranges from macro- to mesoscale dynamics, from large-scale movement of heat and momentum to extreme weather events. The scale of their impact makes understanding storm track dynamics critical to forecasting and climate projections. In this study, we assess CMIP6 historical experiment fidelity to observations of the Southern Hemisphere storm track. Specifically, storm track climatology, variability, and its interactions with low-frequency variability, with the aim of providing confidence for projections of future climate. We find CMIP6 models replicate results from the ERA-5 reanalysis with high fidelity in some regards; namely, capturing climatology of the 500hPa geopotential height field, the role of large-scale variability, and the baroclinic connection with high-frequency variability. However, models fail to capture the magnitude and variability of the storm track, particularly canonical zonal asymmetry. Our results indicate the importance of the storm track is underestimated in CMIP6.

CMIP6 model fidelity in capturing the Southern Hemisphere storm track and its connections with low-frequency variability

I. Campbell[1], J. Renwick[1]

[1]Te Herenga Waka - Victoria, University of Wellington

Key Points:

- CMIP6 models do an excellent job representing mean circulation and low-frequency variability.
- Models simulate the baroclinic connection with storm activity well.
- But models perform poorly in capturing the magnitude of Southern Hemisphere storm activity.

Corresponding author: Isaac Campbell, isaac.campbell6@googlemail.com

Abstract

Storm tracks are a key component of global atmospheric circulation. Their influence ranges from macro- to mesoscale dynamics, from large-scale movement of heat and momentum to extreme weather events. The scale of their impact makes understanding storm track dynamics critical to forecasting and climate projections. In this study, we assess CMIP6 historical experiment fidelity to observations of the Southern Hemisphere storm track. Specifically, storm track climatology, variability, and its interactions with low-frequency variability, with the aim of providing confidence for projections of future climate. We find CMIP6 models replicate results from the ERA-5 reanalysis with high fidelity in some regards; namely, capturing climatology of the 500hPa geopotential height field, the role of large-scale variability, and the baroclinic connection with high-frequency variability. However, models fail to capture the magnitude and variability of the storm track, particularly canonical zonal asymmetry. Our results indicate the importance of the storm track is underestimated in CMIP6.

Plain Language Summary

Storm tracks are regions of considerable storm activity, appearing as a band-like structure around the mid-latitudes. They have a significant role in moving energy and moisture poleward, and are closely associated with extreme weather, such as heavy rainfall, and flooding. Storm tracks tend to vary in strength over time and wander across meridians. Atmospheric circulation phenomena, such as the Southern Annular Mode, interact with the storm track and can influence the location and direction of storms. We look at how well state-of-the-art models simulate the storm track and their interactions with circulation phenomena. We find that models generally do an excellent job characterising the relevant basic circulation, but the strength of the storm track is considerably underestimated. This likely has consequences for the reliability of future climate projections, as it is generally agreed that storm activity is due to increase.

1 Introduction

The Southern Hemisphere Storm Track & Low-Frequency Variability

The Southern Hemisphere (SH) storm track is a region of maximum storm activity covering the Southern Ocean in a band between approximately 40° and 65°S. It is an emergent property of the atmosphere, and a key component of circulation patterns – it has a leading role in the global circulation of momentum, energy, and moisture (Peixoto & Oort, 1992). Storms form downstream of maximum baroclinic instabilities in the time-mean westerly flow (Oort & Peixóto, 1983; Chang & Orlanski, 1993; Trenberth, 1991), and undergo baroclinic and barotropic growth processes (Chang et al., 2002; O'Gorman, 2010). The storm track is closely associated with extremes of wind speed, cloud formation, and precipitation, and greatly impacts weather patterns, including extreme events, through its influence on the behaviour of baroclinic storms (Yettella & Kay, 2017; Pfahl & Wernli, 2012).

Interactions between synoptic storm activity and large-scale low-frequency (LF) variability directly impact global circulation patterns (Kidston et al., 2010; Hoerling & Ting, 1994). These modes of variability manifest as organised spatial patterns of circulation anomalies, driving fluctuating meridional gradients, thus stimulating baroclinicity. This study builds upon previous findings, investigating large-scale circulation patterns that dominate SH circulation variability – in particular, the Southern Annular Mode (SAM), and the El Niño-Southern Oscillation (ENSO) mid-latitude teleconnection – and how these phenomena interact with the storm track.

The SAM is a modal phenomenon with a positive and negative phase (SAM+ and SAM–), manifesting as changes in average circulation over the Antarctic region and mid-latitudes. These circulation anomalies affect storm track position, storm frequency and storm intensity (Lorenz & Hartmann, 2001; Kidston et al., 2010). Close links exist between storm track meridional wandering and the varying meridional pressure gradients that characterise the SAM – it may be said that the SAM essentially defines the meridional location of the storm track. Signatures of SAM+ include reduced average pressures over Antarctica and increased average pressures over the mid-latitudes, resulting in a poleward storm track and strengthening of the jet, whilst the reverse is true under SAM– conditions (Fogt & Marshall, 2020; Hartmann & Lo, 1998).

ENSO is a tropical interannual event which directly affects the shape and position of the storm track via its mid-latitude teleconnection (Timmermann et al., 2018; Hoerling & Ting, 1994). Anomalous convection in the tropics drives the formation of Rossby waves which propagate into the extra-tropics. The propagating wave trains cause organised circulation anomalies far from the source, known as teleconnection patterns, such as the Pacific South American (PSA) pattern (Mo & Higgins, 1998). The storm track in turn amplifies and even controls the propagation of wave trains to preferred locations (Kok & Opsteegh, 1985; Hoerling & Ting, 1994), creating a positive feedback loop (Trenberth et al., 1998; Reboita et al., 2015).

CMIP6 Models

Earth Systems Models (ESMs) are the most complex models contributing to the Intergovernmental Panel on Climate Change (IPCC) Assessment Report 6 (AR6) (Chen et al., 2021). The sixth Coupled Model Intercomparison Project (CMIP6) is the latest in a series of projects which coordinates modelling groups contributing ESMs for a coherent, organised approach within the climate research community. ESMs run time integrations from initial conditions, with realistic mechanics and constraints, to characterise the probability distribution of weather states. Instantaneous states are independent of real-world observations; however, the climatology of each time integration should match observations.

While models generally do an excellent job of simulating the climate system, they are approximations of the real system and often produce mischaracterisations of the global mean state. Part of this error is attributed to model uncertainty, introduced by the construction of a model. Knutti (2018) splits model uncertainty into three sub-categories: model structure; grid resolution; and parameterisation. These are the core representational uncertainties in modelling the climate, belonging solely to model design. Utilising a multi-model ensemble (MME) mean, assuming a degree of model independence, is an effective way to mitigate the influence of model-specific errors, such as those owing to a chosen model tuning strategy (Hourdin et al., 2017). By using many realisations of independent models, it is hoped a wider exploration of the parameter space – an N-dimensional space encompassing all possible model outcomes, where N is the number of degrees of freedom stemming from variables like parameterisation and tuning strategy – will compensate for errors, and produce the expected climate. This is observed by Gleckler et al. (2008), among others, who find the MME mean consistently outperforms any single model.

However, the presence of structural uncertainties undermines projections – all contributing members may possess the same biases that fail to capture a key process or driver of climate change (Parker, 2011, 2013). These systematic errors force the probability distribution of weather states in a consistent way, and are known to impact predictions of an evolving climate. For example, Kidston and Gerber (2010) find biases in the SH jet stream mean state explain inter-model differences in projected trends. This highlights the importance of diagnosing systematic biases, and knowledge of these biases provide a basis for universal model improvement (Flato et al., 2013).

The Storm Track in CMIPs Past & Present

The IPCC 5th Assessment Report (AR5) concluded CMIP5 models capture general characteristics of the storm tracks, albeit with less consistency for the SH. However, most models underestimated storm intensity and frequency (Flato et al., 2013). Bracegirdle et al. (2020) find the storm track equatorward bias has been reduced from 1.9° in CMIP5 to 0.4° in CMIP6. They also find significant improvements in jet variability using decorrelation times – though a positive bias remains – and CMIP6 models successfully simulate a positive SAM trend. Priestley et al. (2020) suggest improvements in CMIP6 models has effectively neutralised the equatorward bias.

Priestley et al. (2020) also find models underestimate peak intensities and bomb cyclone frequency in the SH, which they attribute to a poorly captured intensification process. Chemke et al. (2022) find CMIP6 models do not capture trends in the barotropic growth rate caused by a changing meridional structure of mean zonal winds, specifically the rate of change of meridional gradients – an important driver of eddy growth. They also find models do not capture the momentum convergence around the flanks of storm tracks, as identified in reanalyses by Kidston and Vallis (2010). This, they argue, causes a severe underestimation of the observed SH winter positive eddy kinetic energy (EKE) trend. Whilst a clear signal emerges in the early 21st century in the latest reanalyses, no such signal emerges in models until at least the mid-21st century.

Study Aims

The storm track, SAM and ENSO dominate SH circulation patterns, and forced changes will have a significant impact at the hemispheric and regional scales. The companion paper (Campbell & Renwick, 2023, under review) uses the latest reanalyses and finds the SAM and ENSO teleconnection are important organisers of storm activity. Assessing how well models capture these key features is vital to provide confidence in projections. Changes to the SH storm track characteristics were not explicitly assessed in AR5, and continue to receive less attention. We investigate whether these models successfully capture storm track climatology and variability, and its interactions with LF variability, by assessing the representational accuracy of CMIP6 models. Although representational accuracy does not provide a complete validation of a model, it does provide evidence to support or deny model "fitness-for-purpose", and should be used as part of a wider body of evidence (Knutti, 2018; Parker, 2020).

We compare the climatology, specified by the latest reanalyses, with the historical experiment of an ensemble of 20 CMIP6 models. We characterise model base state, and apply the Common Basis Function (CBF) method to investigate model representation of relevant large-scale variability. Similarly, we derive CBFs of singular vectors, retrieved from a Maximum Covariance Analysis (MCA) on reanalysis data, to establish whether CMIP6 models capture the baroclinic connection between low-frequency and high-frequency variability, and whether large-scale circulation patterns emerge as leading modes of co-variability. Details of our methodology are given in section 2. Results for the ensemble mean are presented in section 3. A discussion and conclusion are provided in section 4.

2 Methodology and Data

We use daily data from the European Centre for Medium-Range Weather Forecasts' (ECMWF) Reanalysis, 5th generation (ERA-5) (Hersbach et al., 2020), and an ensemble of 20 CMIP6 models' historical experiment, taken from the ESGF Node (`https://`

Table 1. CMIP6 models used in the current study, along with their nominal resolutions.

Model Name	Resolution (km)	*Model Name*	Resolution (km)
ACCESS-CM2	250	EC-Earth3-CC	100
ACCESS-ESM1-5	250	FGOALS-f3-L	100
BCC-CSM2-MR	100	GFDL-CM4	100
CanESM5	500	GFDL-ESM4	100
CESM2	100	HadGEM3-GC31-MM	100
CESM2-WACCM	100	MPI-ESM-1-2-HAM	250
CMCC-CM2-SR5	100	NESM3	250
CMCC-ESM2	100	NorCPM1	250
EC-Earth3	100	NorESM2-MM	100
EC-Earth3-AerChem	100	SAM0-UNICON	100

`esgf-node.llnl.gov/search/cmip6/`). We also analysed the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017), and found results are qualitatively very similar to ERA-5; we therefore use only ERA-5 as the reference reanalysis for comparison. It is assumed the chosen sub-ensemble is representative of the wider CMIP6 ensemble. Following the findings of Lee et al. (2019), we assume inter-model variability is of greater importance, and robust inferences can be drawn from CMIP comparisons without considering internal variability. Therefore, we use a single realisation from each model.

To investigate the effect of storm activity on the general circulation, we adopt an Eulerian method, using the 500hPa geopotential height (Z500) field, and its associated HF variance. We use deseasonalised monthly mean Z500 anomalies to which characterise the influence of LF variability, and monthly HF variance anomalies to quantify the effect of passing storm centres. Variance due to storm activity is isolated with a 2–8-day bandpass filter, as per Trenberth (1991). 1° latitude-longitude resolution is used for ERA-5 data – adequately resolving synoptic-scale extratropical storms.

Many models do not match ERA-5 resolution; however, baroclinic storms have a typical scale of O(1000km) – much greater than all model nominal resolutions. We assume an insensitivity of the Eulerian method to resolution, as an extension of the findings of Rohrer et al. (2020), with the proviso that nominal resolution is smaller than the synoptic-scale. Therefore, we interpolate model output to a finer grid of 1° latitude-longitude resolution to match ERA-5. Ensemble models and their nominal resolutions are found in Table 1. CMIP6 models generally simulate up to 2014, therefore, we use the 1972-2014 period – in contrast to the 1979-2021 period used for ERA-5 – to maintain a constant sample size and result robustness. The same analysis was conducted on the 1979-2014 period and the results are qualitatively similar; therefore, results for 1972-2014 are presented below.

We evaluate Z500 and HF variance climatologies and temporal variability by calculating the difference from ERA-5 data to establish biases, for time-mean fields and temporal standard deviation (SD). We use temporally and zonally averaged fields to identify the mean storm track position in two ways: as the maximum meridional Z500 gradient; and as the meridional HF variance maximum. A Savitzky-Golay filter is applied to smooth artefacts from spectral cores (Savitzky & Golay, 1964) – a sensitivity test revealed a window length of 20 is stable. Peak positions are interpolated from smoothed data.

Seasonal Taylor diagrams reveal model ability to capture spatial variability. These diagrams summarise four statistical quantities, namely spatial correlation, R, centred root mean square difference (RMSD), E', and the spatial SDs of the target (model) field, σ_f, and of a reference (ERA-5) field, σ_r. The original paper provides a full description of the diagram, and the relationship between the four quantities (Taylor, 2001). For a simpler presentation, allowing both mean Z500 and HF variance fields to be plotted on the same diagram, E' and σ_f are normalised by σ_r.

We apply the CBF method, outlined by Lee et al. (2019), to assess model fidelity in recreating large-scale variability. Briefly, this method regresses model output onto a reference EOF, derived from ERA-5 in this case, thereby generating an analogous PC, referred to as a CBF-PC. A model equivalent pattern is reconstructed by linearly regressing the CBF-PC onto anomalies at each grid cell, and resultant coefficients are scaled by the CBF-PC SD. Explained variance is defined as the ratio between the area-weighted temporal variance of a reconstructed spatio-temporal field and that of the full field. This method provides a consistent approach allowing for direct comparison with the observational dataset.

Large-scale variability assessed includes: the SAM, characterised by the first EOF of the Z500 deseasonalised anomalies south of 20°S; the Pacific South American patterns, PSA1 and PSA2, characterised by the second and third EOFs of the same Z500 anomalies, as per Fogt and Bromwich (2006); the asymmetrical component of the SAM (A-SAM), characterised by the leading EOF of the Z500 deseasonalised zonal anomalies – with the zonal-mean removed – south of 20°S, following the method of Campitelli et al. (2022); and Zonal Wavenumber-3 (ZW3), characterised by the first and second EOFs of meridional wind anomalies between 40°S and 70°S, as per Goyal et al. (2022).

Similarly, to identify whether CMIP6 models capture the connections identified in Campbell and Renwick (2023, under review), we apply the CBF method to singular vectors specified by modes retrieved from an MCA applied to ERA-5. An MCA performs a singular value decomposition on the cross-covariance matrix of two fields, and extracts singular vectors in order of importance. Details of the MCA are given in Campbell and Renwick (2023, under review), and references therein. Spatial correlation between CBFs and their observational equivalents are calculated to quantify fidelity. A nine-point weighted smoothing function is applied to CBFs to remove residual spectral effects that obscure broad-scale patterns.

3 Results

3.1 Base State

3.1.1 Storm Track Meridional Position

Previous CMIP generations possessed a poleward bias in mean meridional position of the SH storm track (Kidston & Gerber, 2010); Priestley et al. (2020) and Bracegirdle et al. (2020) find this is almost neutralised in CMIP6. To verify these findings, we define the storm track as the peak Z500 meridional gradient, and peak meridional HF variance. Peak positions are provided in Table 2, along with meridional profiles in Figure 1. Some spectral effects are still evident in the profiles despite smoothing, exhibiting noise-like variation.

Generally, all ensemble members identify the reanalysis peak position to within 2° latitude. The ensemble mean meridional peak is only 0.3° equatorward of the ERA-5 climatological position for the zonal-mean Z500 meridional gradient peak, and the zonal-mean HF variance peak matches observations exactly. The ensemble mean Z500 meridional gradient profile is virtually indistinguishable from the ERA-5 profile, despite differing peak positions. The shape of the ensemble mean HF variance profile is also highly

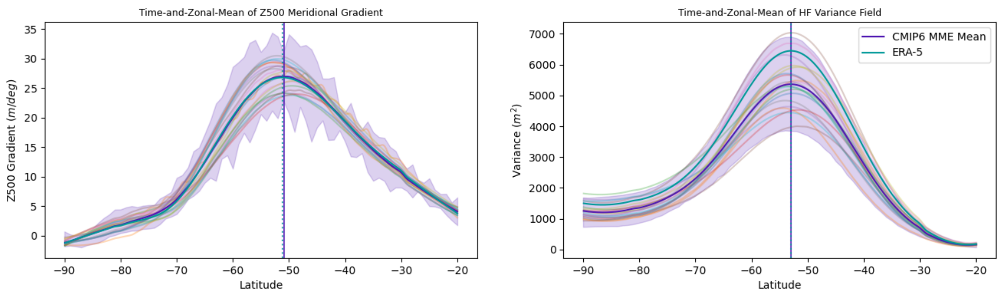

Figure 1. Meridional profiles for mean Z500 gradient and HF variance time-and-zonal mean fields. Vertical lines mark the ensemble mean and ERA-5 peak positions.

similar to ERA-5, but there is a considerable difference in peak amplitude, indicating maximum zonal-mean HF variance is underestimated.

The variance peak is consistently poleward of the Z500 gradient peak in both models and ERA-5. This is perhaps indicative of zonal wind anomalies poleward drift reported by Lorenz and Hartmann (2001). Broadly, if a model exhibits a bias in one peak, so too does the other, suggesting the origin of the bias is linked to both fields. The lack of statistical significance in the difference between ERA-5 peaks and the ensemble mean peak

Table 2. Zonal-and-time-mean meridional peak positions of Z500 meridional gradient and HF variance for all CMIP6 models.

	Z500 Meridional Gradient (°S)	HF Variance (°S)
ERA-5	***51.2***	***53.0***
MME Mean	*50.9*	*53.0*
ACCESS-CM2	49.6	52.4
ACCESS-ESM1-5	50.0	52.6
BCC-CSM2-MR	51.8	52.9
CanESM5	50.0	52.0
CESM2	52.1	53.4
CESM2-WACCM	52.1	53.3
CMCC-CM2-SR5	52.8	53.8
CMCC-ESM2	52.5	53.5
EC-Earth3	50.7	53.4
EC-Earth3-AerChem	49.4	53.4
EC-Earth3-CC	51.0	53.7
FGOALS-f3-L	49.5	52.5
GFDL-CM4	49.0	51.7
GFDL-ESM4	50.3	52.7
HadGEM3-GC31-MM	48.5	51.6
MPI-ESM-1-2-HAM	48.2	51.8
NESM3	51.9	53.5
NorCPM1	53.1	54.4
NorESM2-MM	52.5	54.0
SAM0-UNICON	52.6	54.1

agrees with Priestley et al. (2020) – the bias in CMIP6 is largely neutralised. Bracegirdle et al. (2020) identify a 0.6° equatorward bias, whereas we find a 0.3° equatorward bias in one peak and no bias in the other. This difference is probably due to a difference in the definition of the storm track meridonal position – they use the Jet Latitude Index (JLI) defined by peak zonal-mean winds.

3.1.2 Spatial Variability and Seasonality

Taylor diagrams encapsulate a model's ability to represent spatial patterns and variability of a given field. We present Taylor diagrams for seasonal data of both Z500 and HF variance climatological fields (Figure 2). Each model performs well overall for both fields, with high correlations and relatively low E' in each case. Performance is weaker in the HF variance field; however, pattern correlation remains above 0.95 in all cases. Whereas no model scored below 0.99 in the Z500 field, and SDs are clustered around the reference value, suggesting the climatological Z500 field is well-captured across seasons. Models are considerably weaker in representing HF variance spatial variability. Excepting some models in spring, variability is universally underestimated, with some σ_f almost half that of ERA-5 (see Figure 2). There is also considerable spread of HF variance SDs.

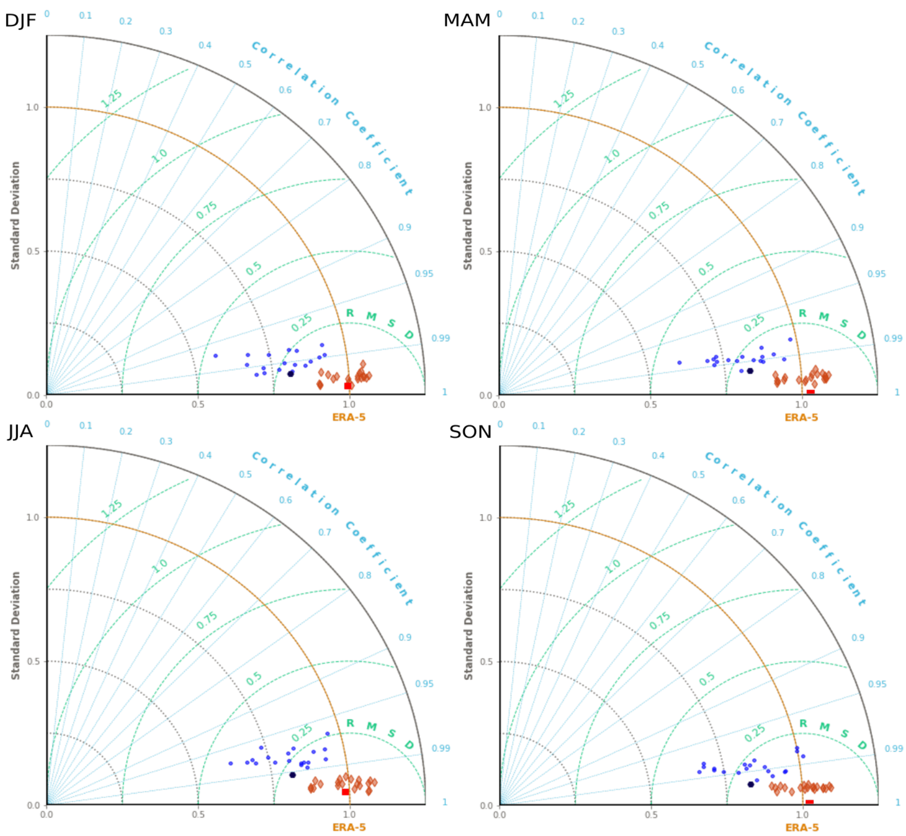

Figure 2. Taylor diagrams for seasonal data. The HF variance field is shown in blue, the mean Z500 field in red, with the MME mean emboldened. Correlation between the model and the reference field is given by the azimuthal angle, field standard deviation, σ_f, by radial distance from the origin, and centered RMSE is proportional to the distance from the reference point (labelled ERA-5). Both SDs and centred RMSEs are presented in normalised units. The yellow arc traces the surface of σ_f equal to 1.

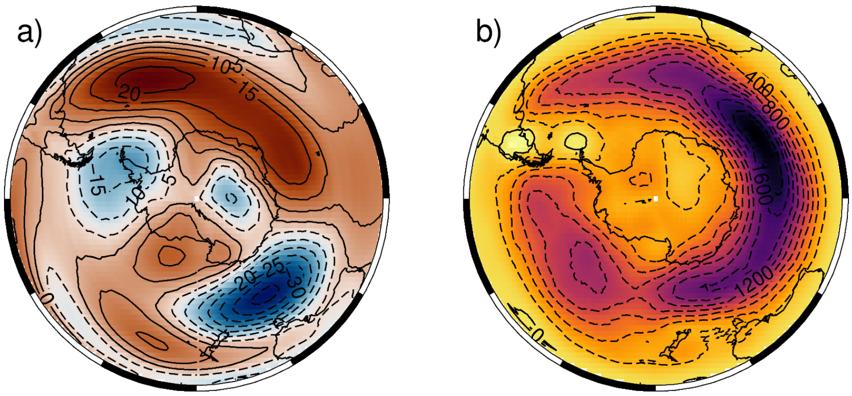

Figure 3. Spatial biases of the multi-model ensemble mean for the a) mean Z500 (m) and b) high-frequency variance (m^2) time-mean fields, relative to the ERA-5 reanalyses.

The ensemble mean is particularly strong in representing aspects of Z500 climatology, with a σ_f approximately equal to ERA-5, and near-perfect correlation in all seasons. The MME mean possesses a stronger correlation for HF variance than most ensemble members. Whereas σ_f is affected by the generally poor ensemble performance, yet retains 82% of the ERA-5 spatial variability. This trend is reflected across seasons. There is no clear seasonal variation in performance, indicating relative performance is consistent across seasons.

3.1.3 Mean Field Spatial Biases

To garner insight into regional differences, model biases from the ERA-5 climatological fields and temporal variability are derived, and the ensemble mean biases presented in Figures 3 and 4. Positive values indicate a higher model value (positive bias), and negative values indicate a lower model value (negative bias). Biases for ensemble members are not shown, but some details are discussed.

Most models possess maximum biases on the order of $100m$ in the mean height field, and $3000m^2$ in the HF variance field, although models of better-than-median performance have biases around $50m$ and $2000m^2$. Recurring biases are found in the Z500 field of many ensemble members, such as a positive bias over the mid-Atlantic Ocean, and a negative bias off the south coast of Australia. These common biases persist in the ensemble mean (Figure 3a) – although with weaker amplitude than any ensemble member – indicating these are systematic biases. The bias south of Australia is located approximately between two oppositely-signed temporal SD biases, seen in Figure 4. Whether these are connected is unclear. The upstream positive bias would suggest models have greater fluctuating Z500 fields in this region, whereas the downstream negative bias, about 30% of the size, would imply less variation in the Z500 field. This is slightly upstream of the Amundsen Sea region (ASR), a particularly active region for propagating Rossby waves – its possible there may be some link.

Models display a near universal bias toward weaker HF variance over the expected storm track position. This is strongest over the Atlantic and Indian oceans, where ERA-5 results indicate the storm track is strongest (see Figure 1 in Campbell and Renwick (2023, under review)). Model consensus translates into the MME mean (Figure 3b), where the Indian Ocean bias is substantial, around $2000m^2$. There is a weaker HF variance absolute bias over the Pacific; however, this coincides with weaker storm activity. HF variance magnitude in the Pacific is around 14% less than ERA-5, 22% less in the Indian Ocean, indicating storm activity is underestimated in CMIP6 on a hemispheric scale. Along-

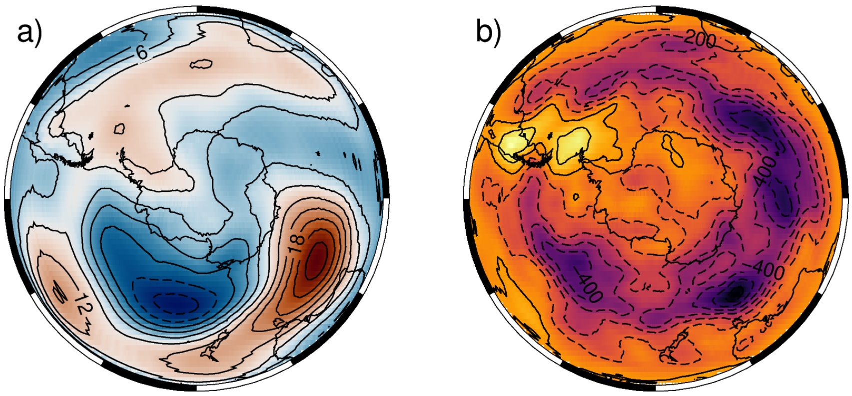

Figure 4. As with Figure 3, but for temporal standard deviations. The temporal SDs are calculated using the monthly mean of each field, over the sample periods of the reanalyses (1979-2021) and CMIP6 models (1972-2014).

side this, Figure 4b shows HF variance temporal variability of the storm track is also weaker, implying the magnitude of passing storms is underestimated. Taken together, this strongly suggests the magnitude of the storm track is widely underestimated in CMIP6.

3.2 Common Basis Functions

3.2.1 Large-scale Modes of Variability

To investigate model fidelity in representing large-scale variability pertinent to the storm track, we reconstruct spatial patterns by regressing model anomalies onto a reference EOF to generate CBFs. The modes of variability studied include the A-SAM – identified by Zhang et al. (2022) to be an area of weakness in CMIP6 – the PSA1 and PSA2 patterns – closely related to the ENSO teleconnection – and ZW3, which plays a dominant role in SH circulation patterns and is expected to have a role in guiding baroclinic eddies. Figure 5 presents the CBFs for the ensemble mean and the reference EOFs from the ERA-5 reanalysis. The MME mean displays high fidelity for all phenomena, as does each model, indicating improvements have been made from the CMIP5 generation (Lee et al., 2019). Ensemble CBFs are provided as supplementary material (Figures S1-S6, supplementary material).

For the SAM, spatial correlation is high for all ensemble members, with a minimum of 0.96 (Figure S1, supplementary material). The ensemble mean has near perfect correlation at 0.99, and low E'. Some models tend to exaggerate the spatial coherence in the annulus surrounding the central polar region, with greater variance in the Pacific. The Atlantic mid-latitude maximum appears to be displaced eastward, which, combined with the more zonally symmetric polar region – both identifiable in the MME mean in Figure 5 – indicates models exaggerate zonal symmetry, or perhaps a summer-like pattern dominates. Those CBFs with a central region protruding over the ASR, similar to ERA-5, tend to have higher correlations.

Zhang et al. (2022) find the A-SAM is poorly represented by CMIP6 models. Upon inspection of the ensemble SAM CBFs (Figure S1, supplementary material), the configuration over the ASR appears quite variable – some appear overly symmetrical, others with exaggerated asymmetry. However, this trend is not identified in the A-SAM CBFs, and correlations and RMSEs are universally strong (Figure S2, supplementary material), translating into a pattern correlation of 0.99 for the ensemble mean (Figure 5). We inspect only annual data, whilst Zhang et al. (2022) subset into seasons. However, they

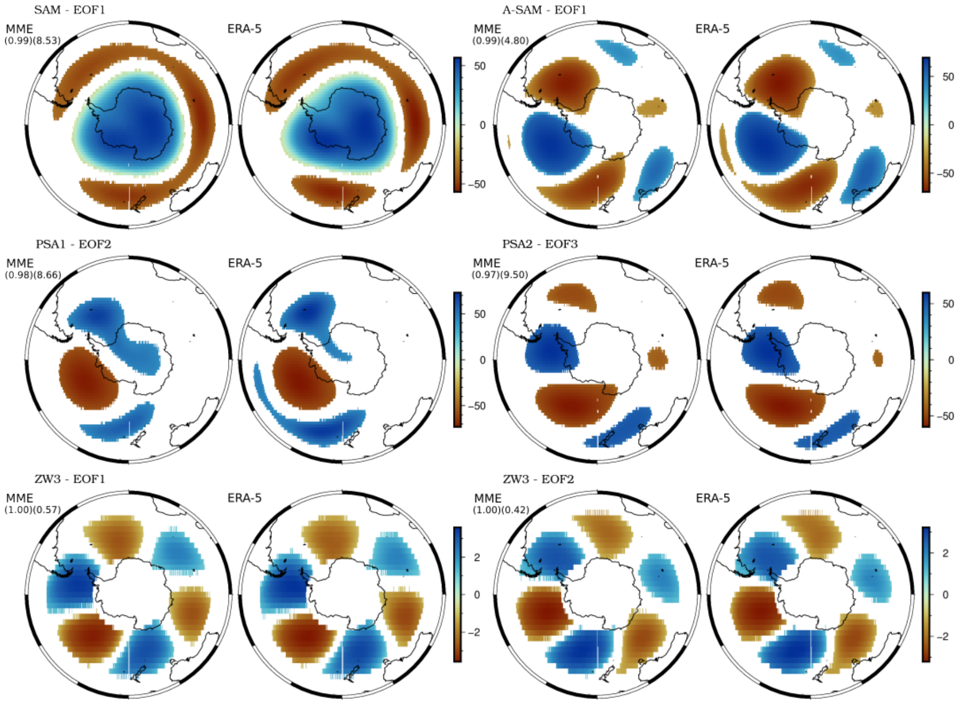

Figure 5. MME mean CBFs characterising circulation variability modes, paired with the original ERA-5 EOFs for reference. Spatial correlation and centred RMSE between the two are provided. Values below a specified magnitude are masked for: the SAM ($|10m|$); A-SAM ($|10m|$); PSA1 ($|15m|$); PSA2 ($|15m|$); ZW3 EOFs ($|0.5ms^{-1}|$).

apply an EOF analysis to each model and find the ensemble mean amplitude is considerably weaker, with comparatively lower pattern correlations. Maximum A-SAM magnitude for both ERA-5 and the ensemble mean, derived through the CBF method, is comparable, 69.3m and 67.4m, respectively. Whether the CBF method finds similar results across seasonal data should be the subject of a future study.

The PSA1 and PSA2 patterns are well simulated, with high correlations and low RMSE, although the spread is greater than SAM and A-SAM CBFs. There is some variability in the location and extent of the wave train extrema in the ensemble CBFs (Figures S3 and S4, supplementary material). This may be linked to several causes, including differences in source location. Quasi-stationary wave activity is less geographically locked in the SH, due to the minimal land mass, thus allowing for greater variability. This effect is smoothed over in the MME mean, whose extrema are effectively collocated with the reference EOF, retrieving correlations of 0.98 and 0.97 for PSA1 and PSA2, respectively. The EOFs related to ZW3 are simulated with similarly high fidelity, with the ensemble mean possessing a correlation of 1.00 to 2 s.f.

3.2.2 Connections Between LF and HF Variability

Leading annual CBFs of the mean Z500 and HF variance fields indicate CMIP6 models are universally strong in capturing the dominant role of the SAM. Pattern correlation for each ensemble member is above 0.80 in both fields, although the Z500 CBFs generally record higher correlation. This can be seen in the ensemble mean in Figure 6 (annual mode 1), with correlations of 0.87 and 0.82 for the Z500 and HF variance CBFs, respectively. Steep Z500 meridional gradients – where contours are densely populated

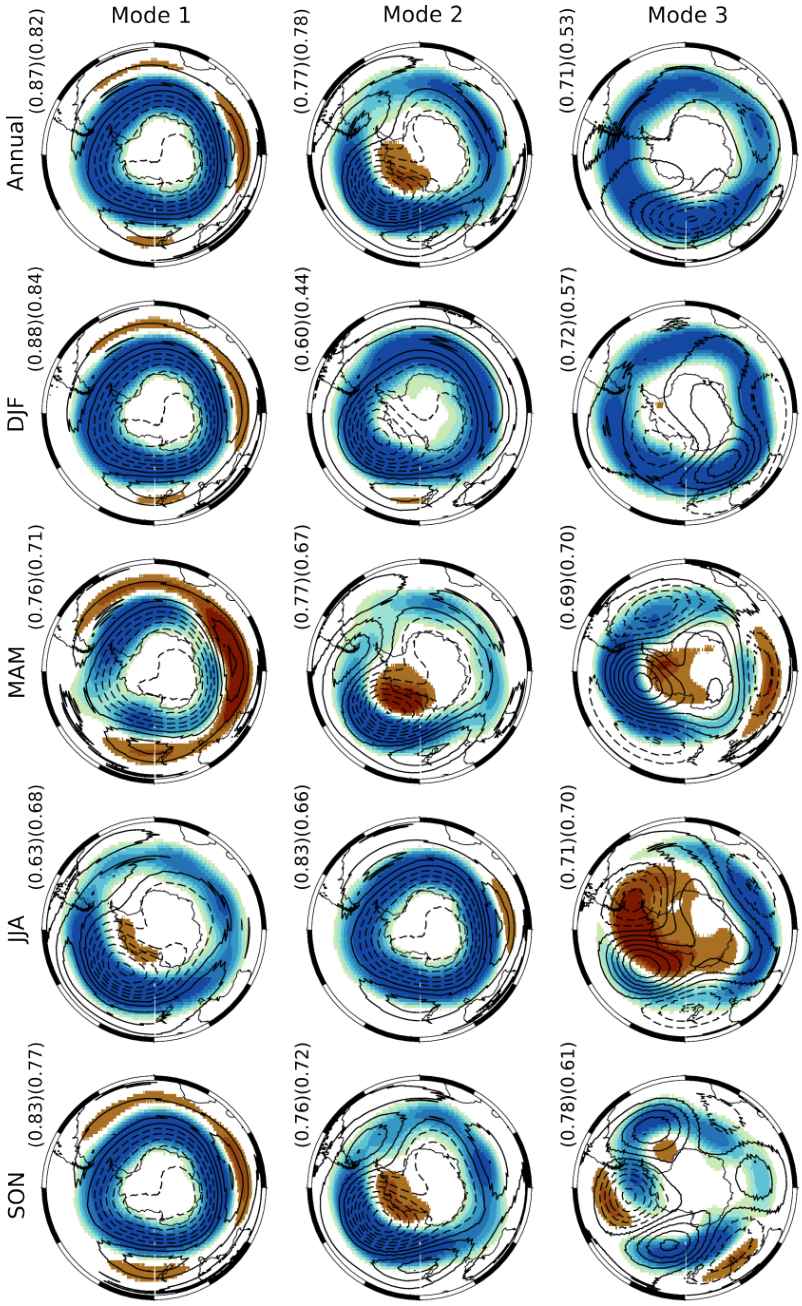

Figure 6. Leading three CBFs of mean Z500 and HF variance fields for the ensemble mean, for annual data and by season, generated from rotated ERA-5 singular vectors. The mean Z500 field is indicated by the contours (10m intervals), positive contours are solid and negative dashed. HF variance is shown as the colour fill, blue indicating increased HF variance and brown decreased variance. HF variance below $|150m^2|$ is masked.

– align with the storm track, suggesting the relationship between baroclinicity and heightened storm activity is well-captured.

This picture is consistent across seasons, with SAM seasonality relatively well-captured in the Z500 field, particularly the varying asymmetrical component. However, the storm track undergoes minimal seasonal variation, and remains broad and coherent even in JJA (see Figure 6, JJA mode 2), when the storm track has a broken appearance according to canonical seasonality (Hoskins & Hodges, 2005). This is expressed by the markedly lower pattern correlation in seasons where storm track asymmetry is pronounced: 0.71 and 0.66 in MAM and JJA, respectively. Note that the ENSO teleconnection- and SAM-like modes are reversed in order in JJA, in accordance with the retrieved modes in Figure 2 in Campbell and Renwick (2023, under review).

The ENSO teleconnection in the second CBFs is similarly well-captured, with the broad-scale pattern present in all ensemble members (Figure S7, supplementary material). This pattern – a wave train pattern across the Pacific with a low centred over the ASR, and leading and trailing highs – persists in the ensemble mean, with high pattern correlations (see Figure 6, annual mode 2). Correlations are weaker than the leading CBFs, though the regional nature of the teleconnection likely contributes. Alignment of the ASL with heightened storm activity in the Pacific mid-latitudes is captured across ensemble members; however, the response of the HF variance field is much wider than ERA-5 suggests, with a weak HF variance pervading the rest of the hemisphere, collocated with the storm track.

Teleconnection seasonality is reasonably well represented in the Z500 field, with a clear SAM-like response in both fields in DJF, and more wave train-like patterns in other seasons. Across the seasons, a ring of HF variance is present, unlike in ERA-5. This ring is particularly strong in DJF, and the correlation drops as low as 0.44. Despite this feature being part of canonical seasonality, it seems to be exaggerated by CMIP6 models, so too is the depth of the ASL.

Ensemble members tend to capture the ZW3 patterns in the third seasonal CBFs (see, for example, SON mode 3 in Figure 6), and pattern correlations remain high. The weaker ZW3 signal in DJF is representative of ERA-5, likely linked to weakened Rossby wave propagation during these months. Note, there is no physical reason for the sign of the HF variance to reverse between seasonal and annual data, but is arbitrarily assigned when singular vectors are retrieved from the ERA-5 MCA. The HF variance ring encircling the pole is visible in all CBFs, and is not found in ERA-5 modes.

The explained variance (EV) of HF variance singular vectors, derived through the CBF method, is fairly stable – around 10% for the first three CBFs – suggesting the importance of HF variance modes of co-variability does not diminish, in direct contrast to ERA-5. On the other hand, the EV of Z500 CBFs decreases from an average 28% in the leading CBF, 16% in the second, and 5% in the third CBF.

4 Conclusion

We have considered output from the CMIP6 programme to assess model performance in representing the SH storm track and associated large-scale variability. We characterised model base state, and applied the CBF method to EOFs and MCA modes to assess fidelity in representing pertinent modes of variability, and connections between the Z500 and HF variance fields. We find CMIP6 models are generally of high fidelity to the ERA-5 reanalysis in capturing Z500 climatology and large-scale variability. The baroclinic connection between the fields – alignment of steep meridional Z500 gradients with increased storm activity – is broadly replicated, so too the organising roles of the SAM and ENSO teleconnection. However, a considerable failure of CMIP6 models is in capturing the magnitude and variability of the storm track, particularly the asymmetrically strong activity in the Indian Ocean. Our results indicate the importance of the storm track is widely underestimated.

The ensemble mean is markedly superior to ensemble members in capturing Z500 climatology, performing at well above the median in all relevant metrics (see Figure 7).

This suggests model-specific errors are compensated for by the ensemble mean. No significant biases were found in storm track position, in agreement with Priestley et al. (2020) and Bracegirdle et al. (2020). Meridional profiles are almost exactly alike, although maximum zonal-mean HF variance is considerably underestimated. CBFs of large-scale variability show universal strong correlation with EOFs derived from the ERA-5 reanalysis, suggesting models successfully capture patterns of LF variability. One caveat with the CBF method is the potential positive bias introduced by using metrics quantifying spatial likeness and variability, when the derivation spatially constrains patterns by regressing onto a reference field, but this effect is not explored in detail here.

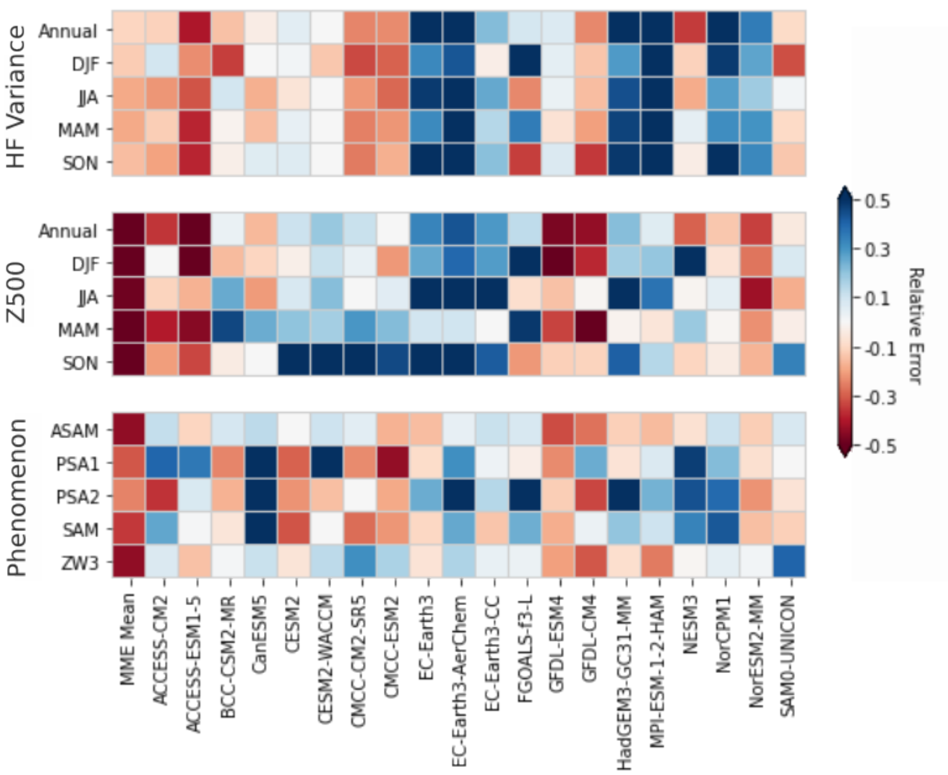

Figure 7. Portrait plot of relative model error for the climatological Z500 and HF variance climatological fields, as well as CBFs for the large-scale variability. Relative error is derived by dividing by median model bias. The error from the two ZW3 CBFs are averaged and presented here as a single value. A negative (red) value indicates better than median performance.

High pattern correlations of leading CBFs suggest all model-derived field anomalies strongly map onto the modes derived from the ERA-5 MCA (Figure 6). This indicates the SAM retains its role in organising HF variability on a hemispheric scale. This result is reinforced by the appearance of the SAM as the leading mode in an MCA applied directly to model output (Figure S8, supplementary material). However, the stable EV in HF variance CBFs suggests its relative importance does not diminish throughout the first three modes. This may be linked to the reduced temporal variability, as seen in Figure 4, perhaps implying storm activity manifests in an overly consistent and symmetrical fashion, in spite of LF variability, and the growth and cessation of passing storms is not well-captured.

The second CBFs (Figures 6 and S7, supplementary material) indicate Rossby wave activity over the Pacific is fairly well-captured. Results from the direct MCA (Figure S8, supplementary material) shows considerable variability in the location of Rossby wave trains over the Pacific; however, considering the strong mapping of model anomalies onto the ERA-5 Z500 singular vectors, this may be a consequence of the dependence of an MCA on sample size, rather than representative of principal locations of the ENSO teleconnection. Given the Pacific sector is a particularly active region for propagating Rossby waves, it is uncertain whether the high pattern correlations in the second CBFs are entirely due to the ENSO teleconnection.

Performance is relatively poor across the ensemble in HF variance spatial and temporal variability, and climatological field biases. This appears to be linked to a hemispheric underestimation of HF variance, particularly over the Indian Ocean. Two models without large HF variance biases, BCC-CSM2-MR and NESM3, are similarly unique in overestimating HF variance spatial variability. Likewise, models with diminished spatial variability are also those with the largest bias in the Indian Ocean, greater than $|4000m^2|$. Why it is that the BCC-CSM2-MR and NESM3 models simulate greater HF variance is unclear, but it seems to have no relation to nominal resolution; this might be the subject of future study.

The zonal asymmetry of the storm track, as described in Hoskins and Hodges (2005), is not well-captured. Figure 2 shows HF variance spatial variability is much lower than the reanalysis, and the asymmetrical biases – a stronger bias in the Indian Ocean relative to the Pacific (Figure 3) – serve to homogenise the storm track, hence smaller spatial SDs, as well as generally underestimating storm activity. Temporal SD biases similarly indicate the amplitude of eddy activity is underestimated. The poor performance in these metrics strongly suggests there is a systematic bias in underrepresenting the role of the storm track, which could have severe consequences for global circulation patterns in ESMs, including the poleward transport of heat and momentum. The impact of this bias on energy transport and circulation patterns will be the subject of a future study.

Typical length-scales of the downstream development process would suggest the biases in the mean Z500 and HF variance fields (Figure 3) are unlikely to be connected, as they are well separated. More likely, the Z500 biases are connected with regional dynamics. The HF variance bias is hemispheric, and likely caused by a failure to capture the relevant physics, such as intensification processes (Priestley et al., 2020). Consistent alignment between steep meridional gradients and increased storm activity indicates baroclinic processes of the storm track are relatively well-captured in CMIP6. Chemke et al. (2022) suggest observations of increasing EKE is due to positive trends in barotropic growth rates caused by changes in meridional zonal-wind structure, which, they find, CMIP6 models fail to capture. The momentum convergence around the flanks, caused by propagating eddies feeding momentum back into the storm track (Lorenz & Hartmann, 2001), is also poorly represented (Chemke et al., 2022), which likely contributes to reduced eddy activity. A critical systematic bias such as this casts considerable doubt as to the validity of storm track projections.

Open Research Section

The ERA-5 data is freely available from the ECMWF Copernicus Online Data Store at `https://cds.climate.copernicus.eu`. MERRA2 data can be found at `https://disc.gsfc.nasa.gov/datasets?project=MERRA-2`. Model output is available at the the ESGF Node (`https://esgf-node.llnl.gov/search/cmip6/`).

Acknowledgments

We would like to thank the authors of the xMCA python library, source code found at `https://github.com/nicrie/xmca`. IC is grateful to the Westergaard family, for their generous provision of Rachael Westergaard's Memorial scholarship. This work was supported by VUW research Support grant 400177.

References

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: 10.1029/2019EA001065

Campbell, I., & Renwick, J. (2023). Southern hemisphere storm tracks and large-scale variability: what do the latest reanalyses say? *Journal of Climate (under review)*.

Campitelli, E., Díaz, L. B., & Vera, C. (2022, January). Assessment of zonally symmetric and asymmetric components of the Southern Annular Mode using a novel approach. *Climate Dynamics*, *58*(1-2), 161–178. doi: 10.1007/s00382-021-05896-5

Chang, E. K. M., Lee, S., & Swanson, K. L. (2002, August). Storm Track Dynamics. *Journal of Climate*, *15*(16), 2163–2183. doi: 10.1175/1520-0442(2002)015⟨02163:STD⟩2.0.CO;2

Chang, E. K. M., & Orlanski, I. (1993, April). On the Dynamics of a Storm Track. *Journal of the Atmospheric Sciences*, *50*(7), 999–1015. doi: 10.1175/1520-0469(1993)050⟨0999:OTDOAS⟩2.0.CO;2

Chemke, R., Ming, Y., & Yuval, J. (2022, May). The intensification of winter mid-latitude storm tracks in the Southern Hemisphere. *Nature Climate Change*, 1–5. doi: 10.1038/s41558-022-01368-8

Chen, D., Rojas, M., Samset, B., Cobb, K. M., Diongue Niang, A., Edwards, P., ... Tréguier, A.-M. (2021). Framing, Context and Methods. In V. Masson-Delmotte et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 147–286). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. In T. Stocker et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Fogt, R. L., & Bromwich, D. H. (2006, March). Decadal Variability of the ENSO Teleconnection to the High-Latitude South Pacific Governed by Coupling with the Southern Annular Mode. *Journal of Climate*, *19*(6), 979–997. doi: 10.1175/JCLI3671.1

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: 10.1002/wcc.652

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017, July). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D6). doi: 10.1029/2007JD008972

Goyal, R., Jucker, M., Gupta, A. S., & England, M. H. (2022, August). A New Zonal Wave-3 Index for the Southern Hemisphere. *Journal of Climate*, *35*(15), 5137–5149. doi: 10.1175/JCLI-D-21-0927.1

Hartmann, D. L., & Lo, F. (1998, April). Wave-Driven Zonal Flow Vacillation in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *55*(8), 1303–1315. doi: 10.1175/1520-0469(1998)055⟨1303:WDZFVI⟩2.0.CO;2

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: 10.1002/qj.3803

Hoerling, M. P., & Ting, M. (1994, May). Organization of Extratropical Transients during El Niño. *Journal of Climate*, *7*(5), 745–766. doi: 10.1175/1520-0442(1994)007⟨0745:OOETDE⟩2.0.CO;2

Hoskins, B. J., & Hodges, K. I. (2005, October). A New Perspective on Southern Hemisphere Storm Tracks. *Journal of Climate*, *18*(20), 4108–4129. doi: 10.1175/JCLI3570.1

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017, March). The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. doi: 10.1175/BAMS-D-15-00135.1

Kidston, J., Frierson, D. M. W., Renwick, J. A., & Vallis, G. K. (2010, December). Observations, Simulations, and Dynamics of Jet Stream Variability and Annular Modes. *Journal of Climate*, *23*(23), 6186–6199. doi: 10.1175/2010JCLI3235.1

Kidston, J., & Gerber, E. P. (2010). Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophysical Research Letters*, *37*(9), L09708. doi: 10.1029/2010GL042873

Kidston, J., & Vallis, G. K. (2010). Relationship between eddy-driven jet latitude and width. *Geophysical Research Letters*, *37*(21), L21809. doi: 10.1029/2010GL044849

Knutti, R. (2018). Climate Model Confirmation: From Philosophy to Predicting Climate in the Real World. In E. A. Lloyd & E. Winsberg (Eds.), *Climate Modelling: Philosophical and Conceptual Issues* (pp. 325–359). Cham: Springer International Publishing.

Kok, C. J., & Opsteegh, J. D. (1985, April). Possible Causes of Anomalies in Seasonal Mean Circulation Patterns during the 1982–83 El Niño Event. *Journal of the Atmospheric Sciences*, *42*(7), 677–694. doi: 10.1175/1520-0469(1985)042⟨0677:PCOAIS⟩2.0.CO;2

Lee, J., Sperber, K. R., Gleckler, P. J., Bonfils, C. J. W., & Taylor, K. E. (2019, April). Quantifying the agreement between observed and simulated extratropical modes of interannual variability. *Climate Dynamics*, *52*(7), 4057–4089. doi: 10.1007/s00382-018-4355-4

Lorenz, D. J., & Hartmann, D. L. (2001, November). Eddy–Zonal Flow Feedback in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *58*(21), 3312–3327. doi: 10.1175/1520-0469(2001)058⟨3312:EZFFIT⟩2.0.CO;2

Mo, K. C., & Higgins, R. W. (1998, June). The Pacific–South American Modes and Tropical Convection during the Southern Hemisphere Winter. *Monthly Weather Review*, *126*(6), 1581–1596. doi: 10.1175/1520-0493(1998)126⟨1581:TPSAMA⟩2.0.CO;2

Oort, A. H., & Peixóto, J. P. (1983, January). Global Angular Momentum and Energy Balance Requirements from Observations. In B. Saltzman (Ed.), *Advances in Geophysics* (Vol. 25, pp. 355–490). Elsevier.

O'Gorman, P. A. (2010, November). Understanding the varied response of the extratropical storm tracks to climate change. *Proceedings of the National Academy of Sciences*, *107*(45), 19176–19180. doi: 10.1073/pnas.1011547107

Parker, W. S. (2011, October). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, *78*(4), 579–600. doi: 10.1086/661566

Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *WIREs Climate Change*, *4*(3), 213–223. doi: 10.1002/wcc.220

Parker, W. S. (2020, July). Model Evaluation: An Adequacy-for-Purpose View. *Philosophy of Science*, *87*(3), 457–477. doi: 10.1086/708691

Peixoto, J. P., & Oort, A. H. (1992). *Physics of climate*. New York, NY, US: American Institute of Physics.

Pfahl, S., & Wernli, H. (2012, October). Quantifying the Relevance of Cyclones for Precipitation Extremes. *Journal of Climate*, *25*(19), 6770–6780. doi: 10.1175/JCLI-D-11-00705.1

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020, August). An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations. *Journal of Climate*, *33*(15), 6315–6343. doi: 10.1175/JCLI-D-19-0928.1

Reboita, M. S., da Rocha, R. P., Ambrizzi, T., & Gouveia, C. D. (2015, October). Trend and teleconnection patterns in the climatology of extratropical cyclones over the Southern Hemisphere. *Climate Dynamics*, *45*(7), 1929–1944. doi: 10.1007/s00382-014-2447-3

Rohrer, M., Martius, O., Raible, C. C., & Brönnimann, S. (2020). Sensitivity of Blocks and Cyclones in ERA5 to Spatial Resolution and Definition. *Geophysical Research Letters*, *47*(7), e2019GL085582. doi: 10.1029/2019GL085582

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627-1639. doi: 10.1021/ac60214a047

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., ... Zhang, X. (2018, July). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. doi: 10.1038/s41586-018-0252-6

Trenberth, K. E. (1991, October). Storm Tracks in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *48*(19), 2159–2178. doi: 10.1175/1520-0469(1991)048⟨2159:STITSH⟩2.0.CO;2

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., & Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, *103*(C7), 14291–14324. doi: 10.1029/97JC01444

Yettella, V., & Kay, J. E. (2017, September). How will precipitation change in extratropical cyclones as the planet warms? Insights from a large initial condition climate model ensemble. *Climate Dynamics*, *49*(5), 1765–1781. doi: 10.1007/s00382-016-3410-2

Zhang, X., He, B., Liu, Y., Bao, Q., Zheng, F., Li, J., ... Wu, G. (2022). Evaluation of the seasonality and spatial aspects of the Southern Annular Mode in CMIP6 models. *International Journal of Climatology*, *42*(7), 3820–3837. doi: 10.1002/joc.7447

CMIP6 model fidelity in capturing the Southern Hemisphere storm track and its connections with low-frequency variability

I. Campbell[1], J. Renwick[1]

[1]Te Herenga Waka - Victoria, University of Wellington

Key Points:

- CMIP6 models do an excellent job representing mean circulation and low-frequency variability.
- Models simulate the baroclinic connection with storm activity well.
- But models perform poorly in capturing the magnitude of Southern Hemisphere storm activity.

Corresponding author: Isaac Campbell, `isaac.campbell6@googlemail.com`

Abstract

Storm tracks are a key component of global atmospheric circulation. Their influence ranges from macro- to mesoscale dynamics, from large-scale movement of heat and momentum to extreme weather events. The scale of their impact makes understanding storm track dynamics critical to forecasting and climate projections. In this study, we assess CMIP6 historical experiment fidelity to observations of the Southern Hemisphere storm track. Specifically, storm track climatology, variability, and its interactions with low-frequency variability, with the aim of providing confidence for projections of future climate. We find CMIP6 models replicate results from the ERA-5 reanalysis with high fidelity in some regards; namely, capturing climatology of the 500hPa geopotential height field, the role of large-scale variability, and the baroclinic connection with high-frequency variability. However, models fail to capture the magnitude and variability of the storm track, particularly canonical zonal asymmetry. Our results indicate the importance of the storm track is underestimated in CMIP6.

Plain Language Summary

Storm tracks are regions of considerable storm activity, appearing as a band-like structure around the mid-latitudes. They have a significant role in moving energy and moisture poleward, and are closely associated with extreme weather, such as heavy rainfall, and flooding. Storm tracks tend to vary in strength over time and wander across meridians. Atmospheric circulation phenomena, such as the Southern Annular Mode, interact with the storm track and can influence the location and direction of storms. We look at how well state-of-the-art models simulate the storm track and their interactions with circulation phenomena. We find that models generally do an excellent job characterising the relevant basic circulation, but the strength of the storm track is considerably underestimated. This likely has consequences for the reliability of future climate projections, as it is generally agreed that storm activity is due to increase.

1 Introduction

The Southern Hemisphere Storm Track & Low-Frequency Variability

The Southern Hemisphere (SH) storm track is a region of maximum storm activity covering the Southern Ocean in a band between approximately 40° and 65°S. It is an emergent property of the atmosphere, and a key component of circulation patterns – it has a leading role in the global circulation of momentum, energy, and moisture (Peixoto & Oort, 1992). Storms form downstream of maximum baroclinic instabilities in the time-mean westerly flow (Oort & Peixóto, 1983; Chang & Orlanski, 1993; Trenberth, 1991), and undergo baroclinic and barotropic growth processes (Chang et al., 2002; O'Gorman, 2010). The storm track is closely associated with extremes of wind speed, cloud formation, and precipitation, and greatly impacts weather patterns, including extreme events, through its influence on the behaviour of baroclinic storms (Yettella & Kay, 2017; Pfahl & Wernli, 2012).

Interactions between synoptic storm activity and large-scale low-frequency (LF) variability directly impact global circulation patterns (Kidston et al., 2010; Hoerling & Ting, 1994). These modes of variability manifest as organised spatial patterns of circulation anomalies, driving fluctuating meridional gradients, thus stimulating baroclinicity. This study builds upon previous findings, investigating large-scale circulation patterns that dominate SH circulation variability – in particular, the Southern Annular Mode (SAM), and the El Niño-Southern Oscillation (ENSO) mid-latitude teleconnection – and how these phenomena interact with the storm track.

The SAM is a modal phenomenon with a positive and negative phase (SAM+ and SAM–), manifesting as changes in average circulation over the Antarctic region and mid-latitudes. These circulation anomalies affect storm track position, storm frequency and storm intensity (Lorenz & Hartmann, 2001; Kidston et al., 2010). Close links exist between storm track meridional wandering and the varying meridional pressure gradients that characterise the SAM – it may be said that the SAM essentially defines the meridional location of the storm track. Signatures of SAM+ include reduced average pressures over Antarctica and increased average pressures over the mid-latitudes, resulting in a poleward storm track and strengthening of the jet, whilst the reverse is true under SAM– conditions (Fogt & Marshall, 2020; Hartmann & Lo, 1998).

ENSO is a tropical interannual event which directly affects the shape and position of the storm track via its mid-latitude teleconnection (Timmermann et al., 2018; Hoerling & Ting, 1994). Anomalous convection in the tropics drives the formation of Rossby waves which propagate into the extra-tropics. The propagating wave trains cause organised circulation anomalies far from the source, known as teleconnection patterns, such as the Pacific South American (PSA) pattern (Mo & Higgins, 1998). The storm track in turn amplifies and even controls the propagation of wave trains to preferred locations (Kok & Opsteegh, 1985; Hoerling & Ting, 1994), creating a positive feedback loop (Trenberth et al., 1998; Reboita et al., 2015).

CMIP6 Models

Earth Systems Models (ESMs) are the most complex models contributing to the Intergovernmental Panel on Climate Change (IPCC) Assessment Report 6 (AR6) (Chen et al., 2021). The sixth Coupled Model Intercomparison Project (CMIP6) is the latest in a series of projects which coordinates modelling groups contributing ESMs for a coherent, organised approach within the climate research community. ESMs run time integrations from initial conditions, with realistic mechanics and constraints, to characterise the probability distribution of weather states. Instantaneous states are independent of real-world observations; however, the climatology of each time integration should match observations.

While models generally do an excellent job of simulating the climate system, they are approximations of the real system and often produce mischaracterisations of the global mean state. Part of this error is attributed to model uncertainty, introduced by the construction of a model. Knutti (2018) splits model uncertainty into three sub-categories: model structure; grid resolution; and parameterisation. These are the core representational uncertainties in modelling the climate, belonging solely to model design. Utilising a multi-model ensemble (MME) mean, assuming a degree of model independence, is an effective way to mitigate the influence of model-specific errors, such as those owing to a chosen model tuning strategy (Hourdin et al., 2017). By using many realisations of independent models, it is hoped a wider exploration of the parameter space – an N-dimensional space encompassing all possible model outcomes, where N is the number of degrees of freedom stemming from variables like parameterisation and tuning strategy – will compensate for errors, and produce the expected climate. This is observed by Gleckler et al. (2008), among others, who find the MME mean consistently outperforms any single model.

However, the presence of structural uncertainties undermines projections – all contributing members may possess the same biases that fail to capture a key process or driver of climate change (Parker, 2011, 2013). These systematic errors force the probability distribution of weather states in a consistent way, and are known to impact predictions of an evolving climate. For example, Kidston and Gerber (2010) find biases in the SH jet stream mean state explain inter-model differences in projected trends. This highlights the importance of diagnosing systematic biases, and knowledge of these biases provide a basis for universal model improvement (Flato et al., 2013).

The Storm Track in CMIPs Past & Present

The IPCC 5th Assessment Report (AR5) concluded CMIP5 models capture general characteristics of the storm tracks, albeit with less consistency for the SH. However, most models underestimated storm intensity and frequency (Flato et al., 2013). Bracegirdle et al. (2020) find the storm track equatorward bias has been reduced from 1.9° in CMIP5 to 0.4° in CMIP6. They also find significant improvements in jet variability using decorrelation times – though a positive bias remains – and CMIP6 models successfully simulate a positive SAM trend. Priestley et al. (2020) suggest improvements in CMIP6 models has effectively neutralised the equatorward bias.

Priestley et al. (2020) also find models underestimate peak intensities and bomb cyclone frequency in the SH, which they attribute to a poorly captured intensification process. Chemke et al. (2022) find CMIP6 models do not capture trends in the barotropic growth rate caused by a changing meridional structure of mean zonal winds, specifically the rate of change of meridional gradients – an important driver of eddy growth. They also find models do not capture the momentum convergence around the flanks of storm tracks, as identified in reanalyses by Kidston and Vallis (2010). This, they argue, causes a severe underestimation of the observed SH winter positive eddy kinetic energy (EKE) trend. Whilst a clear signal emerges in the early 21st century in the latest reanalyses, no such signal emerges in models until at least the mid-21st century.

Study Aims

The storm track, SAM and ENSO dominate SH circulation patterns, and forced changes will have a significant impact at the hemispheric and regional scales. The companion paper (Campbell & Renwick, 2023, under review) uses the latest reanalyses and finds the SAM and ENSO teleconnection are important organisers of storm activity. Assessing how well models capture these key features is vital to provide confidence in projections. Changes to the SH storm track characteristics were not explicitly assessed in AR5, and continue to receive less attention. We investigate whether these models successfully capture storm track climatology and variability, and its interactions with LF variability, by assessing the representational accuracy of CMIP6 models. Although representational accuracy does not provide a complete validation of a model, it does provide evidence to support or deny model "fitness-for-purpose", and should be used as part of a wider body of evidence (Knutti, 2018; Parker, 2020).

We compare the climatology, specified by the latest reanalyses, with the historical experiment of an ensemble of 20 CMIP6 models. We characterise model base state, and apply the Common Basis Function (CBF) method to investigate model representation of relevant large-scale variability. Similarly, we derive CBFs of singular vectors, retrieved from a Maximum Covariance Analysis (MCA) on reanalysis data, to establish whether CMIP6 models capture the baroclinic connection between low-frequency and high-frequency variability, and whether large-scale circulation patterns emerge as leading modes of co-variability. Details of our methodology are given in section 2. Results for the ensemble mean are presented in section 3. A discussion and conclusion are provided in section 4.

2 Methodology and Data

We use daily data from the European Centre for Medium-Range Weather Forecasts' (ECMWF) Reanalysis, 5th generation (ERA-5) (Hersbach et al., 2020), and an ensemble of 20 CMIP6 models' historical experiment, taken from the ESGF Node (`https://`

Table 1. CMIP6 models used in the current study, along with their nominal resolutions.

Model Name	Resolution (km)	*Model Name*	Resolution (km)
ACCESS-CM2	250	EC-Earth3-CC	100
ACCESS-ESM1-5	250	FGOALS-f3-L	100
BCC-CSM2-MR	100	GFDL-CM4	100
CanESM5	500	GFDL-ESM4	100
CESM2	100	HadGEM3-GC31-MM	100
CESM2-WACCM	100	MPI-ESM-1-2-HAM	250
CMCC-CM2-SR5	100	NESM3	250
CMCC-ESM2	100	NorCPM1	250
EC-Earth3	100	NorESM2-MM	100
EC-Earth3-AerChem	100	SAM0-UNICON	100

`esgf-node.llnl.gov/search/cmip6/`). We also analysed the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) (Gelaro et al., 2017), and found results are qualitatively very similar to ERA-5; we therefore use only ERA-5 as the reference reanalysis for comparison. It is assumed the chosen sub-ensemble is representative of the wider CMIP6 ensemble. Following the findings of Lee et al. (2019), we assume inter-model variability is of greater importance, and robust inferences can be drawn from CMIP comparisons without considering internal variability. Therefore, we use a single realisation from each model.

To investigate the effect of storm activity on the general circulation, we adopt an Eulerian method, using the 500hPa geopotential height (Z500) field, and its associated HF variance. We use deseasonalised monthly mean Z500 anomalies to which characterise the influence of LF variability, and monthly HF variance anomalies to quantify the effect of passing storm centres. Variance due to storm activity is isolated with a 2–8-day bandpass filter, as per Trenberth (1991). 1° latitude-longitude resolution is used for ERA-5 data – adequately resolving synoptic-scale extratropical storms.

Many models do not match ERA-5 resolution; however, baroclinic storms have a typical scale of O(1000km) – much greater than all model nominal resolutions. We assume an insensitivity of the Eulerian method to resolution, as an extension of the findings of Rohrer et al. (2020), with the proviso that nominal resolution is smaller than the synoptic-scale. Therefore, we interpolate model output to a finer grid of 1° latitude-longitude resolution to match ERA-5. Ensemble models and their nominal resolutions are found in Table 1. CMIP6 models generally simulate up to 2014, therefore, we use the 1972-2014 period – in contrast to the 1979-2021 period used for ERA-5 – to maintain a constant sample size and result robustness. The same analysis was conducted on the 1979-2014 period and the results are qualitatively similar; therefore, results for 1972-2014 are presented below.

We evaluate Z500 and HF variance climatologies and temporal variability by calculating the difference from ERA-5 data to establish biases, for time-mean fields and temporal standard deviation (SD). We use temporally and zonally averaged fields to identify the mean storm track position in two ways: as the maximum meridional Z500 gradient; and as the meridional HF variance maximum. A Savitzky-Golay filter is applied to smooth artefacts from spectral cores (Savitzky & Golay, 1964) – a sensitivity test revealed a window length of 20 is stable. Peak positions are interpolated from smoothed data.

Seasonal Taylor diagrams reveal model ability to capture spatial variability. These diagrams summarise four statistical quantities, namely spatial correlation, R, centred root mean square difference (RMSD), E', and the spatial SDs of the target (model) field, σ_f, and of a reference (ERA-5) field, σ_r. The original paper provides a full description of the diagram, and the relationship between the four quantities (Taylor, 2001). For a simpler presentation, allowing both mean Z500 and HF variance fields to be plotted on the same diagram, E' and σ_f are normalised by σ_r.

We apply the CBF method, outlined by Lee et al. (2019), to assess model fidelity in recreating large-scale variability. Briefly, this method regresses model output onto a reference EOF, derived from ERA-5 in this case, thereby generating an analogous PC, referred to as a CBF-PC. A model equivalent pattern is reconstructed by linearly regressing the CBF-PC onto anomalies at each grid cell, and resultant coefficients are scaled by the CBF-PC SD. Explained variance is defined as the ratio between the area-weighted temporal variance of a reconstructed spatio-temporal field and that of the full field. This method provides a consistent approach allowing for direct comparison with the observational dataset.

Large-scale variability assessed includes: the SAM, characterised by the first EOF of the Z500 deseasonalised anomalies south of 20°S; the Pacific South American patterns, PSA1 and PSA2, characterised by the second and third EOFs of the same Z500 anomalies, as per Fogt and Bromwich (2006); the asymmetrical component of the SAM (A-SAM), characterised by the leading EOF of the Z500 deseasonalised zonal anomalies – with the zonal-mean removed – south of 20°S, following the method of Campitelli et al. (2022); and Zonal Wavenumber-3 (ZW3), characterised by the first and second EOFs of meridional wind anomalies between 40°S and 70°S, as per Goyal et al. (2022).

Similarly, to identify whether CMIP6 models capture the connections identified in Campbell and Renwick (2023, under review), we apply the CBF method to singular vectors specified by modes retrieved from an MCA applied to ERA-5. An MCA performs a singular value decomposition on the cross-covariance matrix of two fields, and extracts singular vectors in order of importance. Details of the MCA are given in Campbell and Renwick (2023, under review), and references therein. Spatial correlation between CBFs and their observational equivalents are calculated to quantify fidelity. A nine-point weighted smoothing function is applied to CBFs to remove residual spectral effects that obscure broad-scale patterns.

3 Results

3.1 Base State

3.1.1 Storm Track Meridional Position

Previous CMIP generations possessed a poleward bias in mean meridional position of the SH storm track (Kidston & Gerber, 2010); Priestley et al. (2020) and Bracegirdle et al. (2020) find this is almost neutralised in CMIP6. To verify these findings, we define the storm track as the peak Z500 meridional gradient, and peak meridional HF variance. Peak positions are provided in Table 2, along with meridional profiles in Figure 1. Some spectral effects are still evident in the profiles despite smoothing, exhibiting noise-like variation.

Generally, all ensemble members identify the reanalysis peak position to within 2° latitude. The ensemble mean meridional peak is only 0.3° equatorward of the ERA-5 climatological position for the zonal-mean Z500 meridional gradient peak, and the zonal-mean HF variance peak matches observations exactly. The ensemble mean Z500 meridional gradient profile is virtually indistinguishable from the ERA-5 profile, despite differing peak positions. The shape of the ensemble mean HF variance profile is also highly

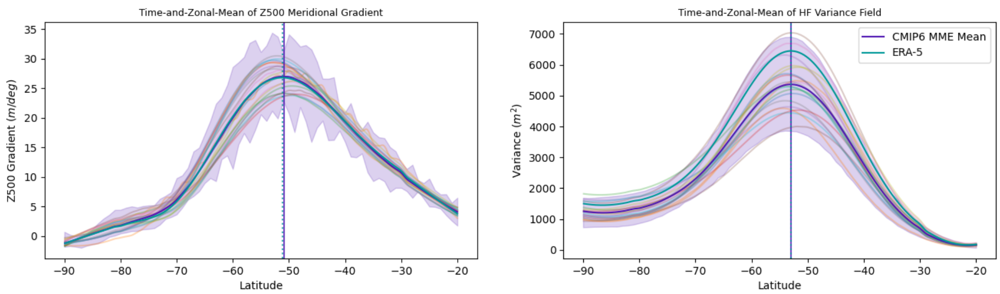

Figure 1. Meridional profiles for mean Z500 gradient and HF variance time-and-zonal mean fields. Vertical lines mark the ensemble mean and ERA-5 peak positions.

similar to ERA-5, but there is a considerable difference in peak amplitude, indicating maximum zonal-mean HF variance is underestimated.

The variance peak is consistently poleward of the Z500 gradient peak in both models and ERA-5. This is perhaps indicative of zonal wind anomalies poleward drift reported by Lorenz and Hartmann (2001). Broadly, if a model exhibits a bias in one peak, so too does the other, suggesting the origin of the bias is linked to both fields. The lack of statistical significance in the difference between ERA-5 peaks and the ensemble mean peak

Table 2. Zonal-and-time-mean meridional peak positions of Z500 meridional gradient and HF variance for all CMIP6 models.

	Z500 Meridional Gradient (°S)	HF Variance (°S)
ERA-5	***51.2***	***53.0***
MME Mean	*50.9*	*53.0*
ACCESS-CM2	49.6	52.4
ACCESS-ESM1-5	50.0	52.6
BCC-CSM2-MR	51.8	52.9
CanESM5	50.0	52.0
CESM2	52.1	53.4
CESM2-WACCM	52.1	53.3
CMCC-CM2-SR5	52.8	53.8
CMCC-ESM2	52.5	53.5
EC-Earth3	50.7	53.4
EC-Earth3-AerChem	49.4	53.4
EC-Earth3-CC	51.0	53.7
FGOALS-f3-L	49.5	52.5
GFDL-CM4	49.0	51.7
GFDL-ESM4	50.3	52.7
HadGEM3-GC31-MM	48.5	51.6
MPI-ESM-1-2-HAM	48.2	51.8
NESM3	51.9	53.5
NorCPM1	53.1	54.4
NorESM2-MM	52.5	54.0
SAM0-UNICON	52.6	54.1

agrees with Priestley et al. (2020) – the bias in CMIP6 is largely neutralised. Bracegirdle et al. (2020) identify a 0.6° equatorward bias, whereas we find a 0.3° equatorward bias in one peak and no bias in the other. This difference is probably due to a difference in the definition of the storm track meridonal position – they use the Jet Latitude Index (JLI) defined by peak zonal-mean winds.

3.1.2 Spatial Variability and Seasonality

Taylor diagrams encapsulate a model's ability to represent spatial patterns and variability of a given field. We present Taylor diagrams for seasonal data of both Z500 and HF variance climatological fields (Figure 2). Each model performs well overall for both fields, with high correlations and relatively low E' in each case. Performance is weaker in the HF variance field; however, pattern correlation remains above 0.95 in all cases. Whereas no model scored below 0.99 in the Z500 field, and SDs are clustered around the reference value, suggesting the climatological Z500 field is well-captured across seasons. Models are considerably weaker in representing HF variance spatial variability. Excepting some models in spring, variability is universally underestimated, with some σ_f almost half that of ERA-5 (see Figure 2). There is also considerable spread of HF variance SDs.

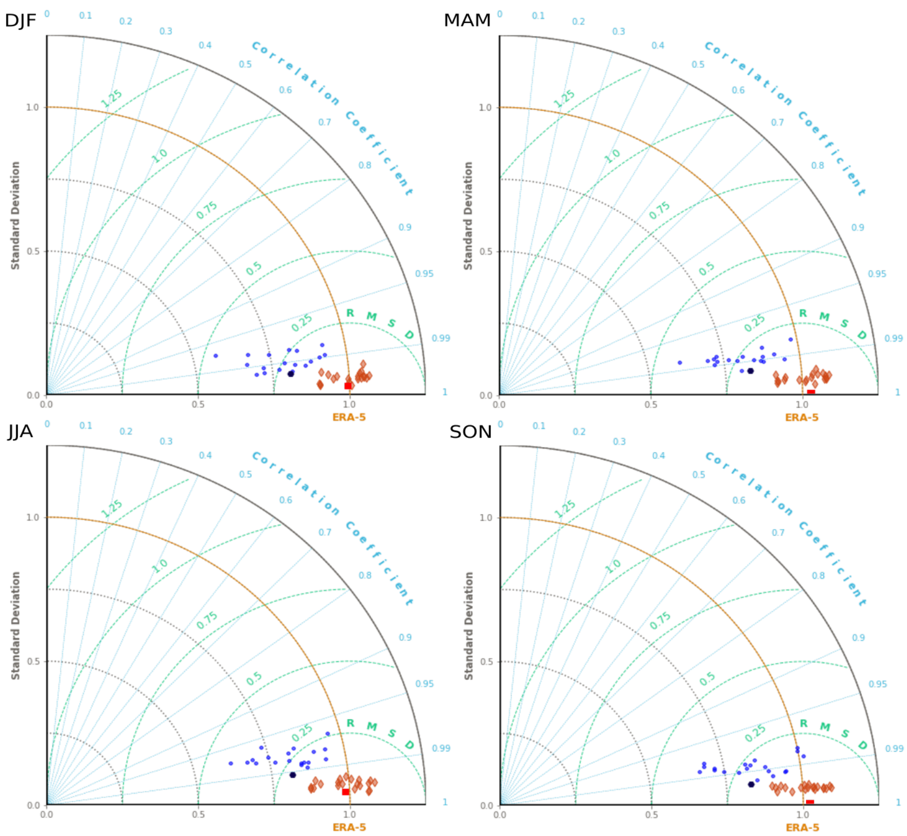

Figure 2. Taylor diagrams for seasonal data. The HF variance field is shown in blue, the mean Z500 field in red, with the MME mean emboldened. Correlation between the model and the reference field is given by the azimuthal angle, field standard deviation, σ_f, by radial distance from the origin, and centered RMSE is proportional to the distance from the reference point (labelled ERA-5). Both SDs and centred RMSEs are presented in normalised units. The yellow arc traces the surface of σ_f equal to 1.

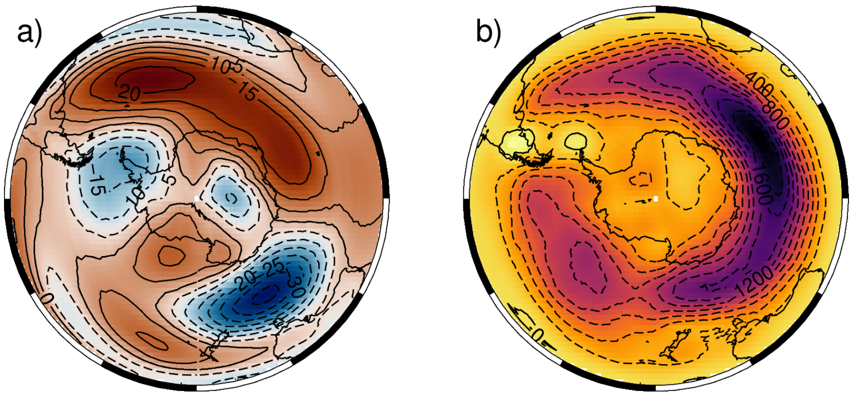

Figure 3. Spatial biases of the multi-model ensemble mean for the a) mean Z500 (m) and b) high-frequency variance (m^2) time-mean fields, relative to the ERA-5 reanalyses.

The ensemble mean is particularly strong in representing aspects of Z500 climatology, with a σ_f approximately equal to ERA-5, and near-perfect correlation in all seasons. The MME mean possesses a stronger correlation for HF variance than most ensemble members. Whereas σ_f is affected by the generally poor ensemble performance, yet retains 82% of the ERA-5 spatial variability. This trend is reflected across seasons. There is no clear seasonal variation in performance, indicating relative performance is consistent across seasons.

3.1.3 Mean Field Spatial Biases

To garner insight into regional differences, model biases from the ERA-5 climatological fields and temporal variability are derived, and the ensemble mean biases presented in Figures 3 and 4. Positive values indicate a higher model value (positive bias), and negative values indicate a lower model value (negative bias). Biases for ensemble members are not shown, but some details are discussed.

Most models possess maximum biases on the order of $100m$ in the mean height field, and $3000m^2$ in the HF variance field, although models of better-than-median performance have biases around $50m$ and $2000m^2$. Recurring biases are found in the Z500 field of many ensemble members, such as a positive bias over the mid-Atlantic Ocean, and a negative bias off the south coast of Australia. These common biases persist in the ensemble mean (Figure 3a) – although with weaker amplitude than any ensemble member – indicating these are systematic biases. The bias south of Australia is located approximately between two oppositely-signed temporal SD biases, seen in Figure 4. Whether these are connected is unclear. The upstream positive bias would suggest models have greater fluctuating Z500 fields in this region, whereas the downstream negative bias, about 30% of the size, would imply less variation in the Z500 field. This is slightly upstream of the Amundsen Sea region (ASR), a particularly active region for propagating Rossby waves – its possible there may be some link.

Models display a near universal bias toward weaker HF variance over the expected storm track position. This is strongest over the Atlantic and Indian oceans, where ERA-5 results indicate the storm track is strongest (see Figure 1 in Campbell and Renwick (2023, under review)). Model consensus translates into the MME mean (Figure 3b), where the Indian Ocean bias is substantial, around $2000m^2$. There is a weaker HF variance absolute bias over the Pacific; however, this coincides with weaker storm activity. HF variance magnitude in the Pacific is around 14% less than ERA-5, 22% less in the Indian Ocean, indicating storm activity is underestimated in CMIP6 on a hemispheric scale. Along-

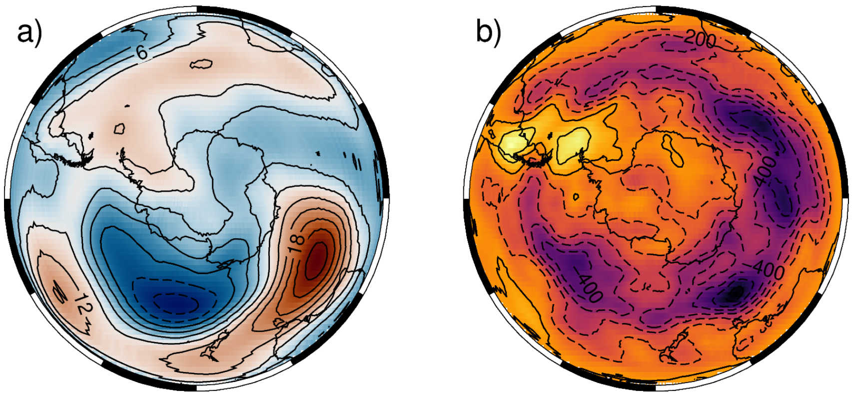

Figure 4. As with Figure 3, but for temporal standard deviations. The temporal SDs are calculated using the monthly mean of each field, over the sample periods of the reanalyses (1979-2021) and CMIP6 models (1972-2014).

side this, Figure 4b shows HF variance temporal variability of the storm track is also weaker, implying the magnitude of passing storms is underestimated. Taken together, this strongly suggests the magnitude of the storm track is widely underestimated in CMIP6.

3.2 Common Basis Functions

3.2.1 Large-scale Modes of Variability

To investigate model fidelity in representing large-scale variability pertinent to the storm track, we reconstruct spatial patterns by regressing model anomalies onto a reference EOF to generate CBFs. The modes of variability studied include the A-SAM – identified by Zhang et al. (2022) to be an area of weakness in CMIP6 – the PSA1 and PSA2 patterns – closely related to the ENSO teleconnection – and ZW3, which plays a dominant role in SH circulation patterns and is expected to have a role in guiding baroclinic eddies. Figure 5 presents the CBFs for the ensemble mean and the reference EOFs from the ERA-5 reanalysis. The MME mean displays high fidelity for all phenomena, as does each model, indicating improvements have been made from the CMIP5 generation (Lee et al., 2019). Ensemble CBFs are provided as supplementary material (Figures S1-S6, supplementary material).

For the SAM, spatial correlation is high for all ensemble members, with a minimum of 0.96 (Figure S1, supplementary material). The ensemble mean has near perfect correlation at 0.99, and low E'. Some models tend to exaggerate the spatial coherence in the annulus surrounding the central polar region, with greater variance in the Pacific. The Atlantic mid-latitude maximum appears to be displaced eastward, which, combined with the more zonally symmetric polar region – both identifiable in the MME mean in Figure 5 – indicates models exaggerate zonal symmetry, or perhaps a summer-like pattern dominates. Those CBFs with a central region protruding over the ASR, similar to ERA-5, tend to have higher correlations.

Zhang et al. (2022) find the A-SAM is poorly represented by CMIP6 models. Upon inspection of the ensemble SAM CBFs (Figure S1, supplementary material), the configuration over the ASR appears quite variable – some appear overly symmetrical, others with exaggerated asymmetry. However, this trend is not identified in the A-SAM CBFs, and correlations and RMSEs are universally strong (Figure S2, supplementary material), translating into a pattern correlation of 0.99 for the ensemble mean (Figure 5). We inspect only annual data, whilst Zhang et al. (2022) subset into seasons. However, they

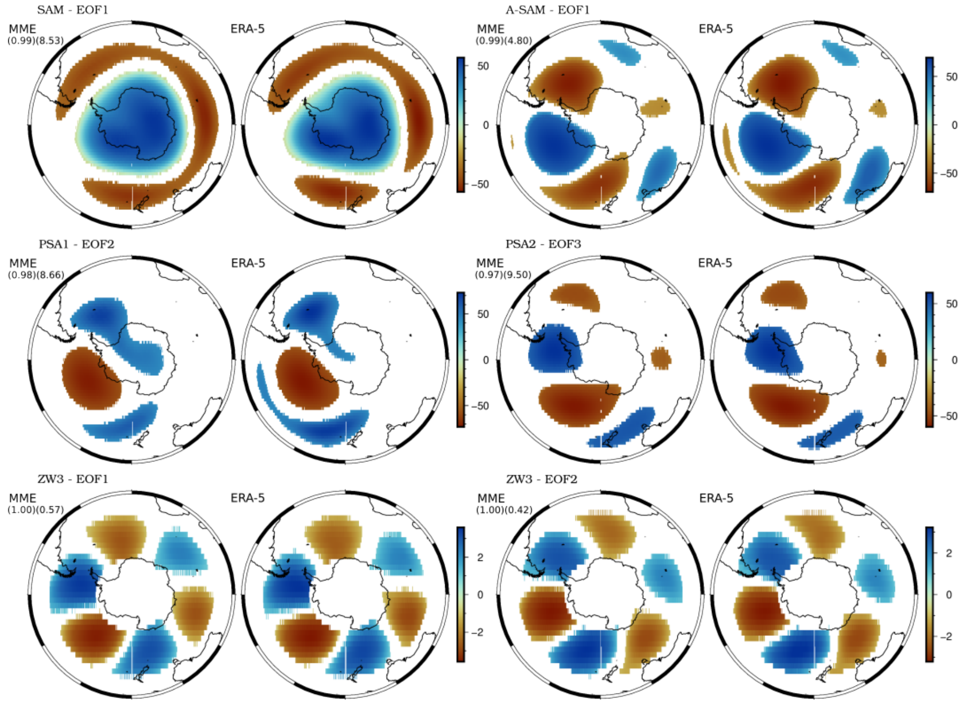

Figure 5. MME mean CBFs characterising circulation variability modes, paired with the original ERA-5 EOFs for reference. Spatial correlation and centred RMSE between the two are provided. Values below a specified magnitude are masked for: the SAM ($|10m|$); A-SAM ($|10m|$); PSA1 ($|15m|$); PSA2 ($|15m|$); ZW3 EOFs ($|0.5ms^{-1}|$).

apply an EOF analysis to each model and find the ensemble mean amplitude is considerably weaker, with comparatively lower pattern correlations. Maximum A-SAM magnitude for both ERA-5 and the ensemble mean, derived through the CBF method, is comparable, 69.3m and 67.4m, respectively. Whether the CBF method finds similar results across seasonal data should be the subject of a future study.

The PSA1 and PSA2 patterns are well simulated, with high correlations and low RMSE, although the spread is greater than SAM and A-SAM CBFs. There is some variability in the location and extent of the wave train extrema in the ensemble CBFs (Figures S3 and S4, supplementary material). This may be linked to several causes, including differences in source location. Quasi-stationary wave activity is less geographically locked in the SH, due to the minimal land mass, thus allowing for greater variability. This effect is smoothed over in the MME mean, whose extrema are effectively collocated with the reference EOF, retrieving correlations of 0.98 and 0.97 for PSA1 and PSA2, respectively. The EOFs related to ZW3 are simulated with similarly high fidelity, with the ensemble mean possessing a correlation of 1.00 to 2 s.f.

3.2.2 Connections Between LF and HF Variability

Leading annual CBFs of the mean Z500 and HF variance fields indicate CMIP6 models are universally strong in capturing the dominant role of the SAM. Pattern correlation for each ensemble member is above 0.80 in both fields, although the Z500 CBFs generally record higher correlation. This can be seen in the ensemble mean in Figure 6 (annual mode 1), with correlations of 0.87 and 0.82 for the Z500 and HF variance CBFs, respectively. Steep Z500 meridional gradients – where contours are densely populated

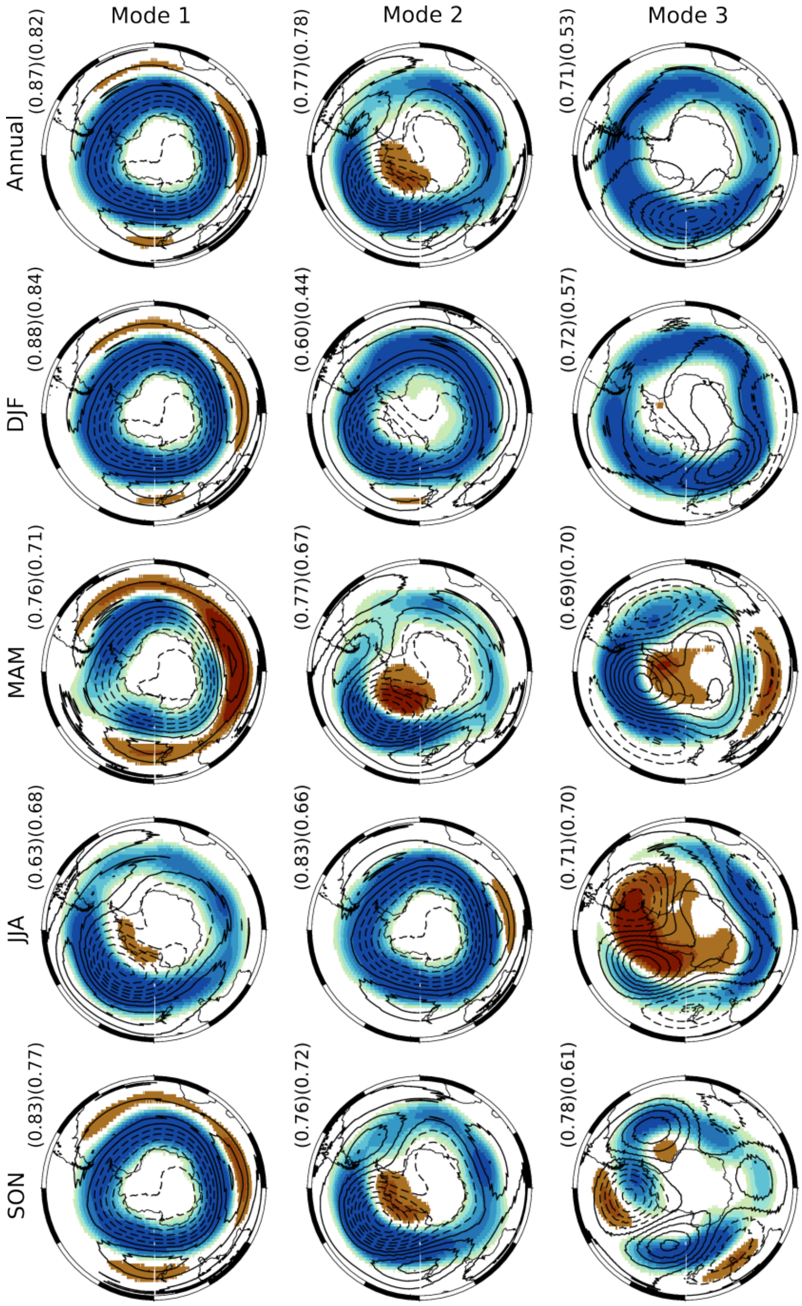

Figure 6. Leading three CBFs of mean Z500 and HF variance fields for the ensemble mean, for annual data and by season, generated from rotated ERA-5 singular vectors. The mean Z500 field is indicated by the contours ($10m$ intervals), positive contours are solid and negative dashed. HF variance is shown as the colour fill, blue indicating increased HF variance and brown decreased variance. HF variance below $|150m^2|$ is masked.

– align with the storm track, suggesting the relationship between baroclinicity and heightened storm activity is well-captured.

This picture is consistent across seasons, with SAM seasonality relatively well-captured in the Z500 field, particularly the varying asymmetrical component. However, the storm track undergoes minimal seasonal variation, and remains broad and coherent even in JJA (see Figure 6, JJA mode 2), when the storm track has a broken appearance according to canonical seasonality (Hoskins & Hodges, 2005). This is expressed by the markedly lower pattern correlation in seasons where storm track asymmetry is pronounced: 0.71 and 0.66 in MAM and JJA, respectively. Note that the ENSO teleconnection- and SAM-like modes are reversed in order in JJA, in accordance with the retrieved modes in Figure 2 in Campbell and Renwick (2023, under review).

The ENSO teleconnection in the second CBFs is similarly well-captured, with the broad-scale pattern present in all ensemble members (Figure S7, supplementary material). This pattern – a wave train pattern across the Pacific with a low centred over the ASR, and leading and trailing highs – persists in the ensemble mean, with high pattern correlations (see Figure 6, annual mode 2). Correlations are weaker than the leading CBFs, though the regional nature of the teleconnection likely contributes. Alignment of the ASL with heightened storm activity in the Pacific mid-latitudes is captured across ensemble members; however, the response of the HF variance field is much wider than ERA-5 suggests, with a weak HF variance pervading the rest of the hemisphere, collocated with the storm track.

Teleconnection seasonality is reasonably well represented in the Z500 field, with a clear SAM-like response in both fields in DJF, and more wave train-like patterns in other seasons. Across the seasons, a ring of HF variance is present, unlike in ERA-5. This ring is particularly strong in DJF, and the correlation drops as low as 0.44. Despite this feature being part of canonical seasonality, it seems to be exaggerated by CMIP6 models, so too is the depth of the ASL.

Ensemble members tend to capture the ZW3 patterns in the third seasonal CBFs (see, for example, SON mode 3 in Figure 6), and pattern correlations remain high. The weaker ZW3 signal in DJF is representative of ERA-5, likely linked to weakened Rossby wave propagation during these months. Note, there is no physical reason for the sign of the HF variance to reverse between seasonal and annual data, but is arbitrarily assigned when singular vectors are retrieved from the ERA-5 MCA. The HF variance ring encircling the pole is visible in all CBFs, and is not found in ERA-5 modes.

The explained variance (EV) of HF variance singular vectors, derived through the CBF method, is fairly stable – around 10% for the first three CBFs – suggesting the importance of HF variance modes of co-variability does not diminish, in direct contrast to ERA-5. On the other hand, the EV of Z500 CBFs decreases from an average 28% in the leading CBF, 16% in the second, and 5% in the third CBF.

4 Conclusion

We have considered output from the CMIP6 programme to assess model performance in representing the SH storm track and associated large-scale variability. We characterised model base state, and applied the CBF method to EOFs and MCA modes to assess fidelity in representing pertinent modes of variability, and connections between the Z500 and HF variance fields. We find CMIP6 models are generally of high fidelity to the ERA-5 reanalysis in capturing Z500 climatology and large-scale variability. The baroclinic connection between the fields – alignment of steep meridional Z500 gradients with increased storm activity – is broadly replicated, so too the organising roles of the SAM and ENSO teleconnection. However, a considerable failure of CMIP6 models is in capturing the magnitude and variability of the storm track, particularly the asymmetrically strong activity in the Indian Ocean. Our results indicate the importance of the storm track is widely underestimated.

The ensemble mean is markedly superior to ensemble members in capturing Z500 climatology, performing at well above the median in all relevant metrics (see Figure 7).

This suggests model-specific errors are compensated for by the ensemble mean. No significant biases were found in storm track position, in agreement with Priestley et al. (2020) and Bracegirdle et al. (2020). Meridional profiles are almost exactly alike, although maximum zonal-mean HF variance is considerably underestimated. CBFs of large-scale variability show universal strong correlation with EOFs derived from the ERA-5 reanalysis, suggesting models successfully capture patterns of LF variability. One caveat with the CBF method is the potential positive bias introduced by using metrics quantifying spatial likeness and variability, when the derivation spatially constrains patterns by regressing onto a reference field, but this effect is not explored in detail here.

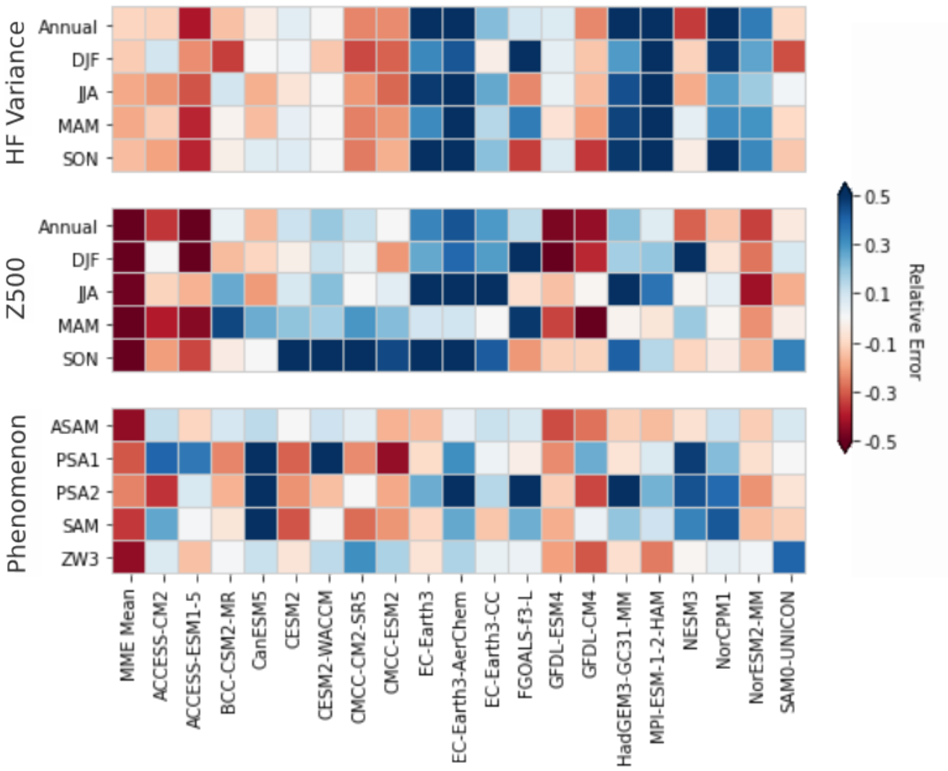

Figure 7. Portrait plot of relative model error for the climatological Z500 and HF variance climatological fields, as well as CBFs for the large-scale variability. Relative error is derived by dividing by median model bias. The error from the two ZW3 CBFs are averaged and presented here as a single value. A negative (red) value indicates better than median performance.

High pattern correlations of leading CBFs suggest all model-derived field anomalies strongly map onto the modes derived from the ERA-5 MCA (Figure 6). This indicates the SAM retains its role in organising HF variability on a hemispheric scale. This result is reinforced by the appearance of the SAM as the leading mode in an MCA applied directly to model output (Figure S8, supplementary material). However, the stable EV in HF variance CBFs suggests its relative importance does not diminish throughout the first three modes. This may be linked to the reduced temporal variability, as seen in Figure 4, perhaps implying storm activity manifests in an overly consistent and symmetrical fashion, in spite of LF variability, and the growth and cessation of passing storms is not well-captured.

The second CBFs (Figures 6 and S7, supplementary material) indicate Rossby wave activity over the Pacific is fairly well-captured. Results from the direct MCA (Figure S8, supplementary material) shows considerable variability in the location of Rossby wave trains over the Pacific; however, considering the strong mapping of model anomalies onto the ERA-5 Z500 singular vectors, this may be a consequence of the dependence of an MCA on sample size, rather than representative of principal locations of the ENSO teleconnection. Given the Pacific sector is a particularly active region for propagating Rossby waves, it is uncertain whether the high pattern correlations in the second CBFs are entirely due to the ENSO teleconnection.

Performance is relatively poor across the ensemble in HF variance spatial and temporal variability, and climatological field biases. This appears to be linked to a hemispheric underestimation of HF variance, particularly over the Indian Ocean. Two models without large HF variance biases, BCC-CSM2-MR and NESM3, are similarly unique in overestimating HF variance spatial variability. Likewise, models with diminished spatial variability are also those with the largest bias in the Indian Ocean, greater than $|4000m^2|$. Why it is that the BCC-CSM2-MR and NESM3 models simulate greater HF variance is unclear, but it seems to have no relation to nominal resolution; this might be the subject of future study.

The zonal asymmetry of the storm track, as described in Hoskins and Hodges (2005), is not well-captured. Figure 2 shows HF variance spatial variability is much lower than the reanalysis, and the asymmetrical biases – a stronger bias in the Indian Ocean relative to the Pacific (Figure 3) – serve to homogenise the storm track, hence smaller spatial SDs, as well as generally underestimating storm activity. Temporal SD biases similarly indicate the amplitude of eddy activity is underestimated. The poor performance in these metrics strongly suggests there is a systematic bias in underrepresenting the role of the storm track, which could have severe consequences for global circulation patterns in ESMs, including the poleward transport of heat and momentum. The impact of this bias on energy transport and circulation patterns will be the subject of a future study.

Typical length-scales of the downstream development process would suggest the biases in the mean Z500 and HF variance fields (Figure 3) are unlikely to be connected, as they are well separated. More likely, the Z500 biases are connected with regional dynamics. The HF variance bias is hemispheric, and likely caused by a failure to capture the relevant physics, such as intensification processes (Priestley et al., 2020). Consistent alignment between steep meridional gradients and increased storm activity indicates baroclinic processes of the storm track are relatively well-captured in CMIP6. Chemke et al. (2022) suggest observations of increasing EKE is due to positive trends in barotropic growth rates caused by changes in meridional zonal-wind structure, which, they find, CMIP6 models fail to capture. The momentum convergence around the flanks, caused by propagating eddies feeding momentum back into the storm track (Lorenz & Hartmann, 2001), is also poorly represented (Chemke et al., 2022), which likely contributes to reduced eddy activity. A critical systematic bias such as this casts considerable doubt as to the validity of storm track projections.

Open Research Section

The ERA-5 data is freely available from the ECMWF Copernicus Online Data Store at `https://cds.climate.copernicus.eu`. MERRA2 data can be found at `https://disc.gsfc.nasa.gov/datasets?project=MERRA-2`. Model output is available at the the ESGF Node (`https://esgf-node.llnl.gov/search/cmip6/`).

Acknowledgments

We would like to thank the authors of the xMCA python library, source code found at `https://github.com/nicrie/xmca`. IC is grateful to the Westergaard family, for their generous provision of Rachael Westergaard's Memorial scholarship. This work was supported by VUW research Support grant 400177.

References

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: 10.1029/2019EA001065

Campbell, I., & Renwick, J. (2023). Southern hemisphere storm tracks and large-scale variability: what do the latest reanalyses say? *Journal of Climate (under review)*.

Campitelli, E., Díaz, L. B., & Vera, C. (2022, January). Assessment of zonally symmetric and asymmetric components of the Southern Annular Mode using a novel approach. *Climate Dynamics*, *58*(1-2), 161–178. doi: 10.1007/s00382-021-05896-5

Chang, E. K. M., Lee, S., & Swanson, K. L. (2002, August). Storm Track Dynamics. *Journal of Climate*, *15*(16), 2163–2183. doi: 10.1175/1520-0442(2002)015⟨02163:STD⟩2.0.CO;2

Chang, E. K. M., & Orlanski, I. (1993, April). On the Dynamics of a Storm Track. *Journal of the Atmospheric Sciences*, *50*(7), 999–1015. doi: 10.1175/1520-0469(1993)050⟨0999:OTDOAS⟩2.0.CO;2

Chemke, R., Ming, Y., & Yuval, J. (2022, May). The intensification of winter mid-latitude storm tracks in the Southern Hemisphere. *Nature Climate Change*, 1–5. doi: 10.1038/s41558-022-01368-8

Chen, D., Rojas, M., Samset, B., Cobb, K. M., Diongue Niang, A., Edwards, P., ... Tréguier, A.-M. (2021). Framing, Context and Methods. In V. Masson-Delmotte et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 147–286). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. In T. Stocker et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Fogt, R. L., & Bromwich, D. H. (2006, March). Decadal Variability of the ENSO Teleconnection to the High-Latitude South Pacific Governed by Coupling with the Southern Annular Mode. *Journal of Climate*, *19*(6), 979–997. doi: 10.1175/JCLI3671.1

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: 10.1002/wcc.652

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017, July). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D6). doi: 10.1029/2007JD008972

Goyal, R., Jucker, M., Gupta, A. S., & England, M. H. (2022, August). A New Zonal Wave-3 Index for the Southern Hemisphere. *Journal of Climate*, *35*(15), 5137–5149. doi: 10.1175/JCLI-D-21-0927.1

Hartmann, D. L., & Lo, F. (1998, April). Wave-Driven Zonal Flow Vacillation in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *55*(8), 1303–1315. doi: 10.1175/1520-0469(1998)055⟨1303:WDZFVI⟩2.0.CO;2

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: 10.1002/qj.3803

Hoerling, M. P., & Ting, M. (1994, May). Organization of Extratropical Transients during El Niño. *Journal of Climate*, *7*(5), 745–766. doi: 10.1175/1520-0442(1994)007⟨0745:OOETDE⟩2.0.CO;2

Hoskins, B. J., & Hodges, K. I. (2005, October). A New Perspective on Southern Hemisphere Storm Tracks. *Journal of Climate*, *18*(20), 4108–4129. doi: 10.1175/JCLI3570.1

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017, March). The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. doi: 10.1175/BAMS-D-15-00135.1

Kidston, J., Frierson, D. M. W., Renwick, J. A., & Vallis, G. K. (2010, December). Observations, Simulations, and Dynamics of Jet Stream Variability and Annular Modes. *Journal of Climate*, *23*(23), 6186–6199. doi: 10.1175/2010JCLI3235.1

Kidston, J., & Gerber, E. P. (2010). Intermodel variability of the poleward shift of the austral jet stream in the CMIP3 integrations linked to biases in 20th century climatology. *Geophysical Research Letters*, *37*(9), L09708. doi: 10.1029/2010GL042873

Kidston, J., & Vallis, G. K. (2010). Relationship between eddy-driven jet latitude and width. *Geophysical Research Letters*, *37*(21), L21809. doi: 10.1029/2010GL044849

Knutti, R. (2018). Climate Model Confirmation: From Philosophy to Predicting Climate in the Real World. In E. A. Lloyd & E. Winsberg (Eds.), *Climate Modelling: Philosophical and Conceptual Issues* (pp. 325–359). Cham: Springer International Publishing.

Kok, C. J., & Opsteegh, J. D. (1985, April). Possible Causes of Anomalies in Seasonal Mean Circulation Patterns during the 1982–83 El Niño Event. *Journal of the Atmospheric Sciences*, *42*(7), 677–694. doi: 10.1175/1520-0469(1985)042⟨0677:PCOAIS⟩2.0.CO;2

Lee, J., Sperber, K. R., Gleckler, P. J., Bonfils, C. J. W., & Taylor, K. E. (2019, April). Quantifying the agreement between observed and simulated extratropical modes of interannual variability. *Climate Dynamics*, *52*(7), 4057–4089. doi: 10.1007/s00382-018-4355-4

Lorenz, D. J., & Hartmann, D. L. (2001, November). Eddy–Zonal Flow Feedback in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *58*(21), 3312–3327. doi: 10.1175/1520-0469(2001)058⟨3312:EZFFIT⟩2.0.CO;2

Mo, K. C., & Higgins, R. W. (1998, June). The Pacific–South American Modes and Tropical Convection during the Southern Hemisphere Winter. *Monthly Weather Review*, *126*(6), 1581–1596. doi: 10.1175/1520-0493(1998)126⟨1581:TPSAMA⟩2.0.CO;2

Oort, A. H., & Peixóto, J. P. (1983, January). Global Angular Momentum and Energy Balance Requirements from Observations. In B. Saltzman (Ed.), *Advances in Geophysics* (Vol. 25, pp. 355–490). Elsevier.

O'Gorman, P. A. (2010, November). Understanding the varied response of the extratropical storm tracks to climate change. *Proceedings of the National Academy of Sciences*, *107*(45), 19176–19180. doi: 10.1073/pnas.1011547107

Parker, W. S. (2011, October). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, *78*(4), 579–600. doi: 10.1086/661566

Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *WIREs Climate Change*, *4*(3), 213–223. doi: 10.1002/wcc.220

Parker, W. S. (2020, July). Model Evaluation: An Adequacy-for-Purpose View. *Philosophy of Science*, *87*(3), 457–477. doi: 10.1086/708691

Peixoto, J. P., & Oort, A. H. (1992). *Physics of climate*. New York, NY, US: American Institute of Physics.

Pfahl, S., & Wernli, H. (2012, October). Quantifying the Relevance of Cyclones for Precipitation Extremes. *Journal of Climate*, *25*(19), 6770–6780. doi: 10.1175/JCLI-D-11-00705.1

Priestley, M. D. K., Ackerley, D., Catto, J. L., Hodges, K. I., McDonald, R. E., & Lee, R. W. (2020, August). An Overview of the Extratropical Storm Tracks in CMIP6 Historical Simulations. *Journal of Climate*, *33*(15), 6315–6343. doi: 10.1175/JCLI-D-19-0928.1

Reboita, M. S., da Rocha, R. P., Ambrizzi, T., & Gouveia, C. D. (2015, October). Trend and teleconnection patterns in the climatology of extratropical cyclones over the Southern Hemisphere. *Climate Dynamics*, *45*(7), 1929–1944. doi: 10.1007/s00382-014-2447-3

Rohrer, M., Martius, O., Raible, C. C., & Brönnimann, S. (2020). Sensitivity of Blocks and Cyclones in ERA5 to Spatial Resolution and Definition. *Geophysical Research Letters*, *47*(7), e2019GL085582. doi: 10.1029/2019GL085582

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627-1639. doi: 10.1021/ac60214a047

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. doi: 10.1029/2000JD900719

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., … Zhang, X. (2018, July). El Niño–Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. doi: 10.1038/s41586-018-0252-6

Trenberth, K. E. (1991, October). Storm Tracks in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *48*(19), 2159–2178. doi: 10.1175/1520-0469(1991)048⟨2159:STITSH⟩2.0.CO;2

Trenberth, K. E., Branstator, G. W., Karoly, D., Kumar, A., Lau, N.-C., & Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *Journal of Geophysical Research: Oceans*, *103*(C7), 14291–14324. doi: 10.1029/97JC01444

Yettella, V., & Kay, J. E. (2017, September). How will precipitation change in extratropical cyclones as the planet warms? Insights from a large initial condition climate model ensemble. *Climate Dynamics*, *49*(5), 1765–1781. doi: 10.1007/s00382-016-3410-2

Zhang, X., He, B., Liu, Y., Bao, Q., Zheng, F., Li, J., … Wu, G. (2022). Evaluation of the seasonality and spatial aspects of the Southern Annular Mode in CMIP6 models. *International Journal of Climatology*, *42*(7), 3820–3837. doi: 10.1002/joc.7447

JOURNAL OF GEOPHYSICAL RESEARCH

Supporting Information for "CMIP6 fidelity in capturing the SH ST"

I. Campbell[1], J. Renwick[1]

[1]Te Herenga Waka - Victoria, University of Wellington

Contents of this file

1. Figures S1 to S8

Introduction

Here, we provide supplementary figures for the article "CMIP6 model fidelity in capturing the Southern Hemisphere storm track and its connections with low-frequency variability". The figures present results for each ensemble member used to calculate the ensemble mean, for annual data only: CBFs of ERA-5 EOFs derived for large-scale modes of variability; model CBFs of ERA-5 MCA modes of Z500 and HF variance monthly anomalies; MCA modes derived directly from CMIP6 model output, for Z500 and HF variance monthly anomalies.

Corresponding author: I. H. Campbell, School of Geography, Environment, and Earth Sciences, Te Herenga Waka - Victoria, University of Wellington, Cotton Building, Wellington, 6012, NZ. (isaac.campbell6@googlemail.com)

1. Ensemble Common Basis Functions

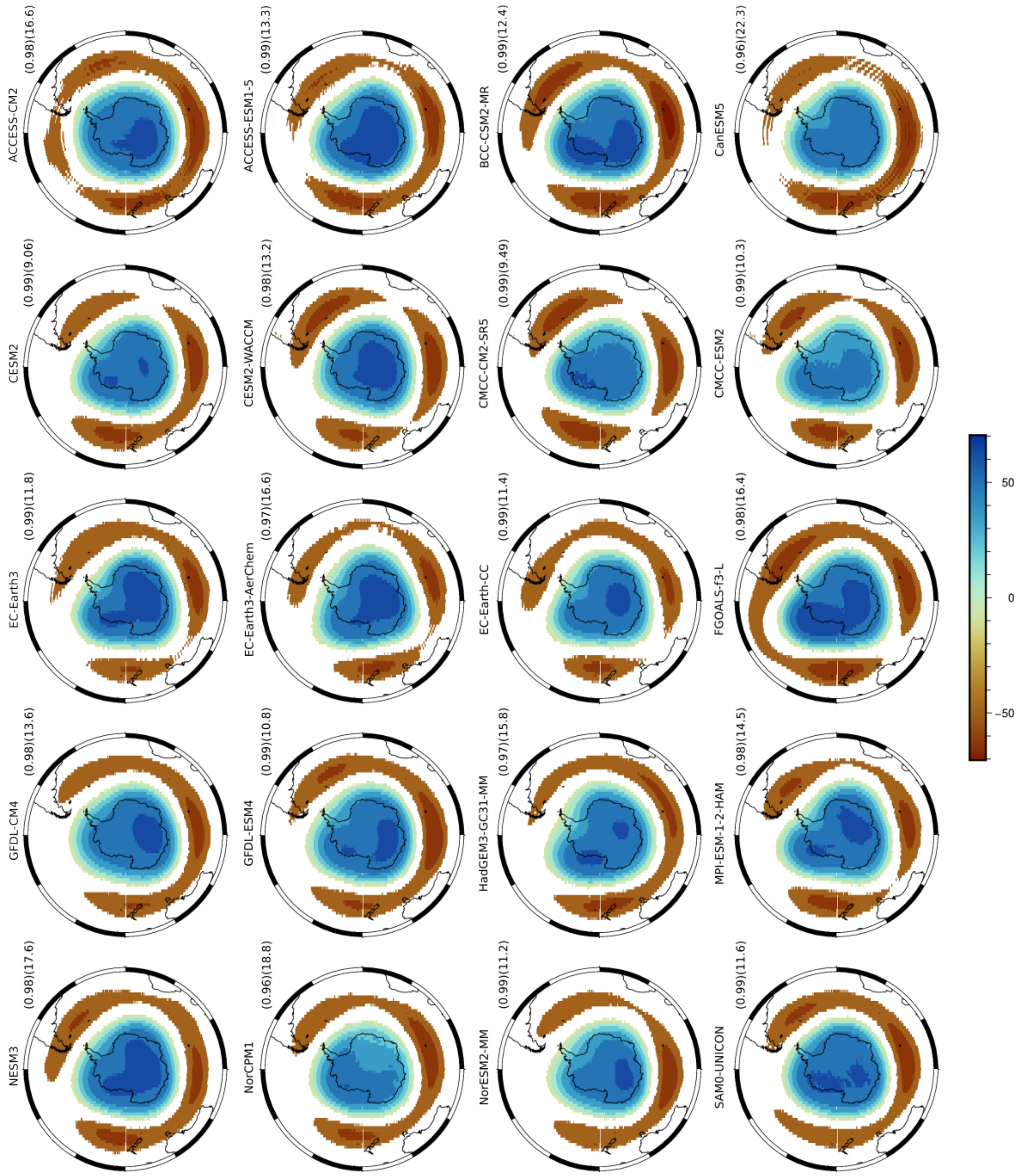

Figure S1. CBFs characterising the Southern Annular Mode (SAM) for all CMIP6 ensemble members. Presented here as colour fill, values below a magnitude of $|10m|$ are supressed.

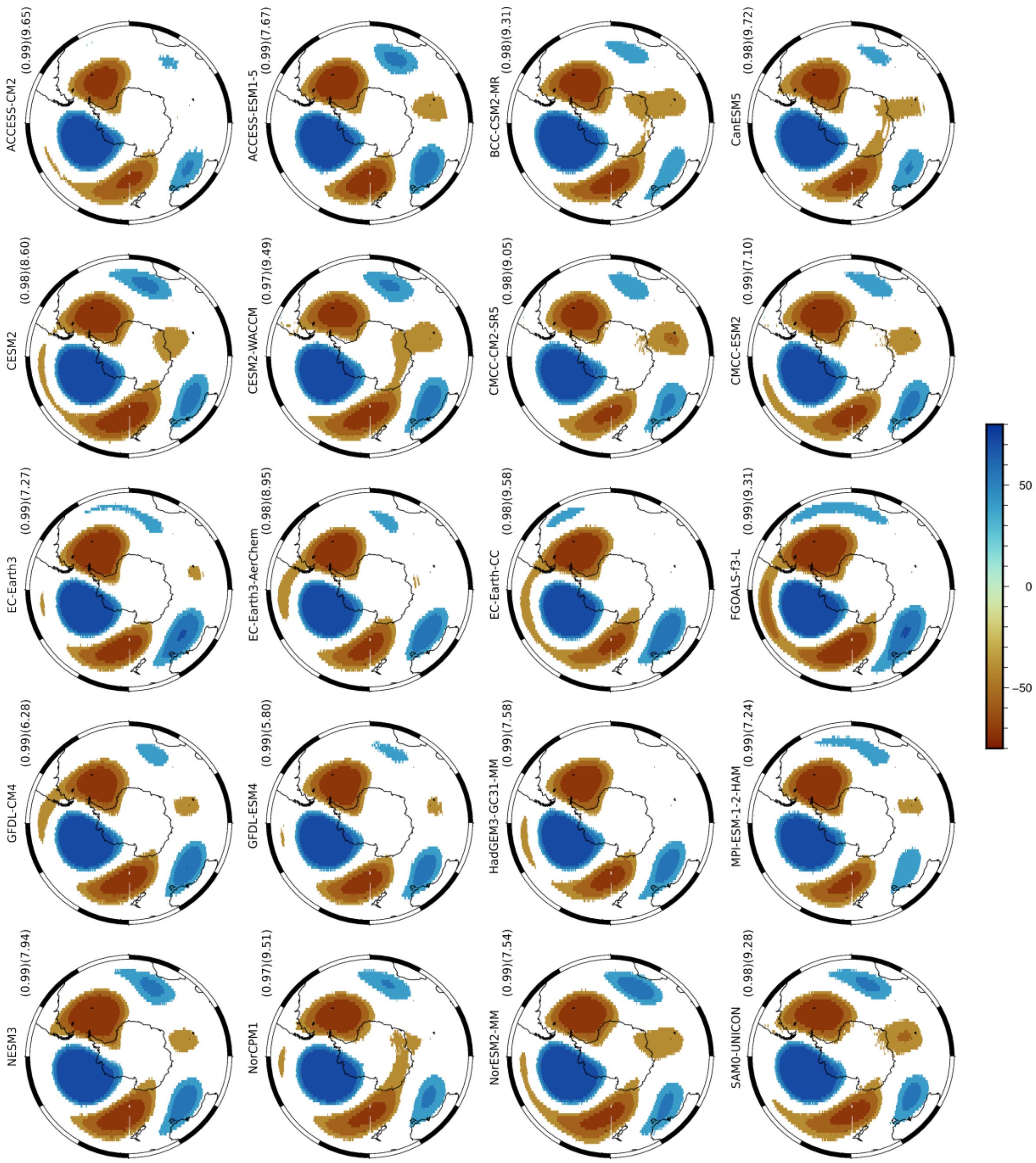

Figure S2. As with Figure S1, but for the asymmetrical SAM, values below a magnitude of $|10m|$ are supressed.

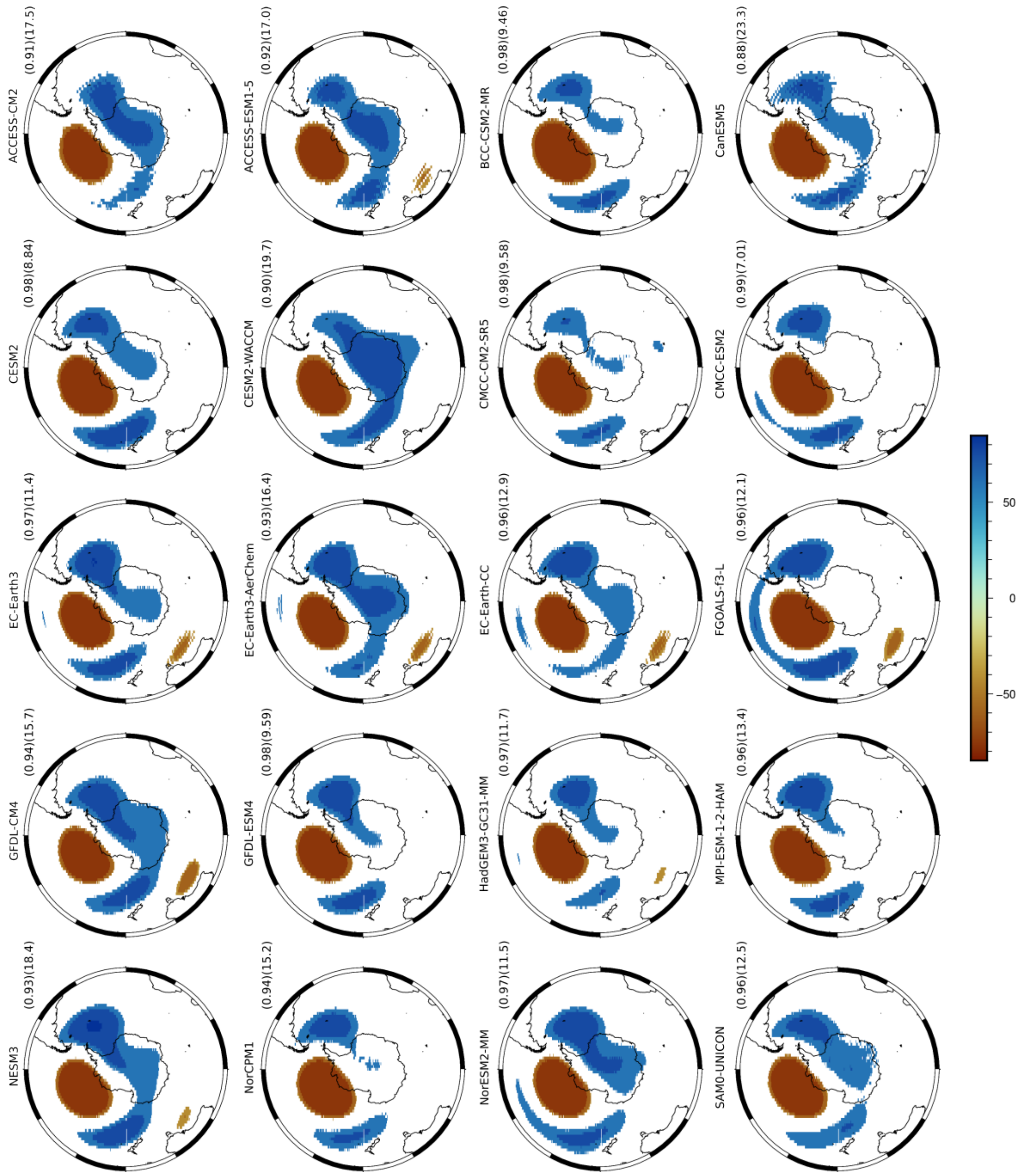

Figure S3. As with Figure S1, but for the Pacific South American pattern 1; values below a magnitude of $|15m|$ are supressed.

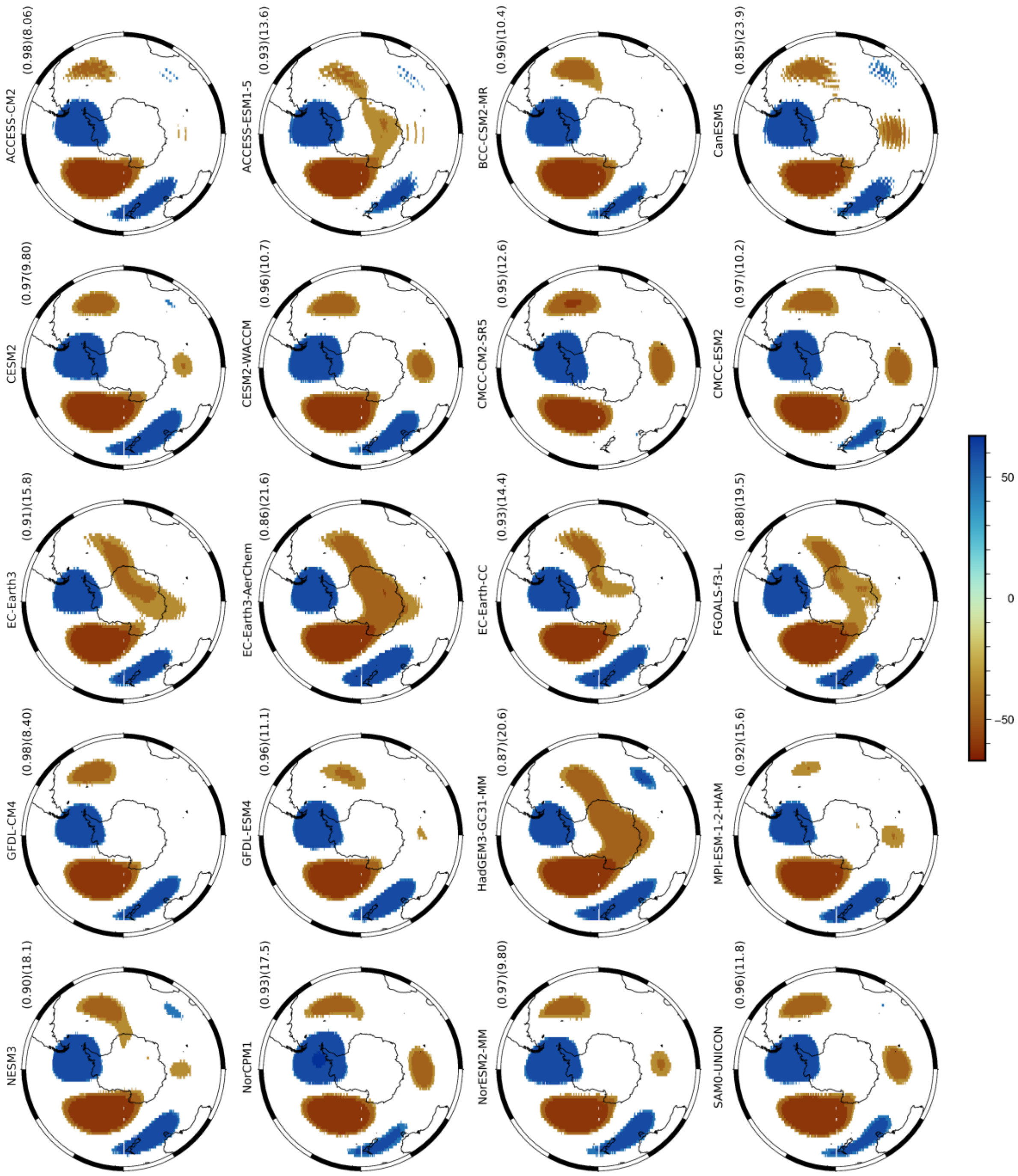

Figure S4. As with Figure S1, but for the Pacific South American pattern 2; values below a magnitude of $|15m|$ are supressed.

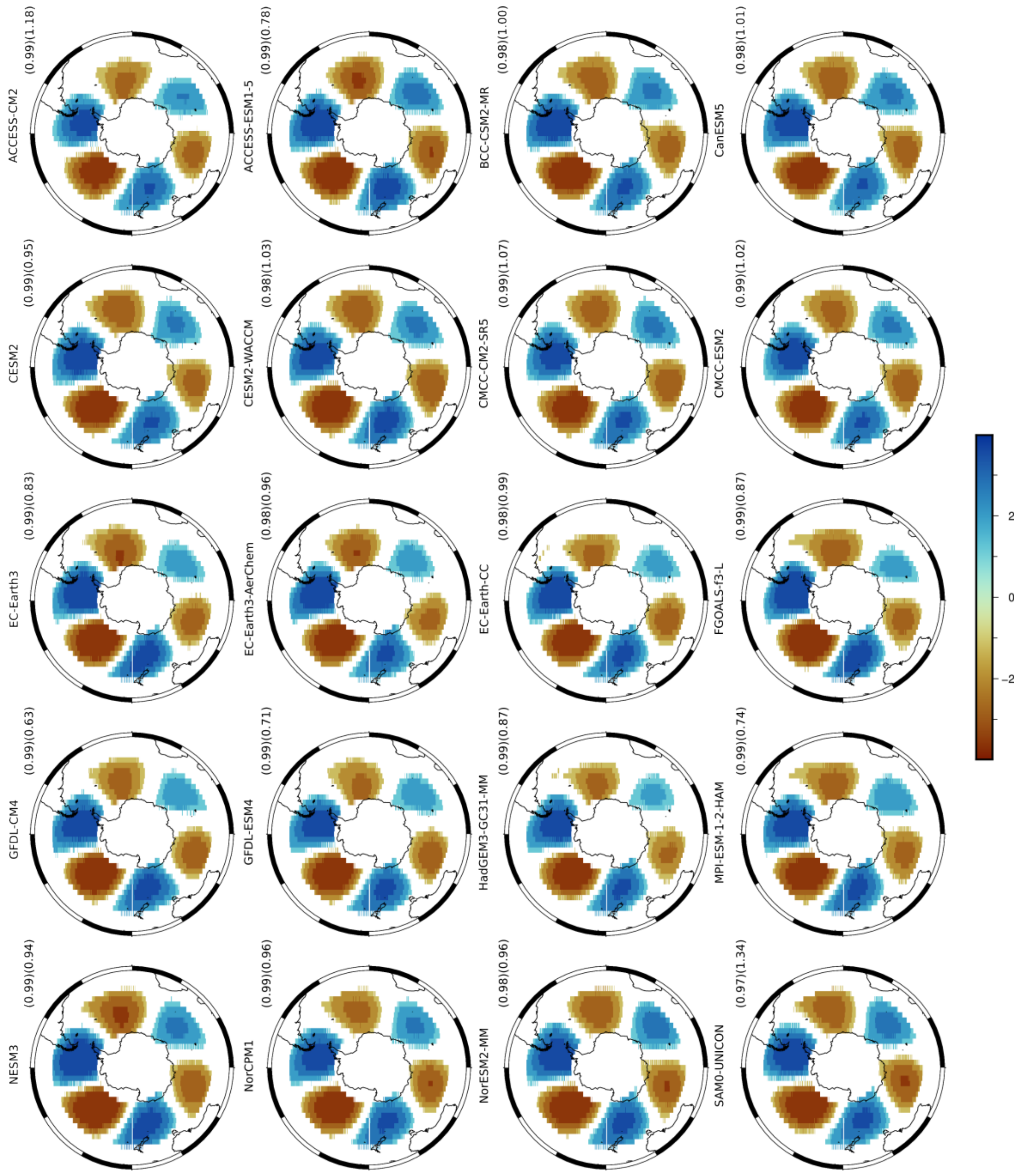

Figure S5. As with Figure S1, but for the Zonal Wave 3, first EOF, values below a magnitude of $|0.5ms^{-1}|$ are supressed.

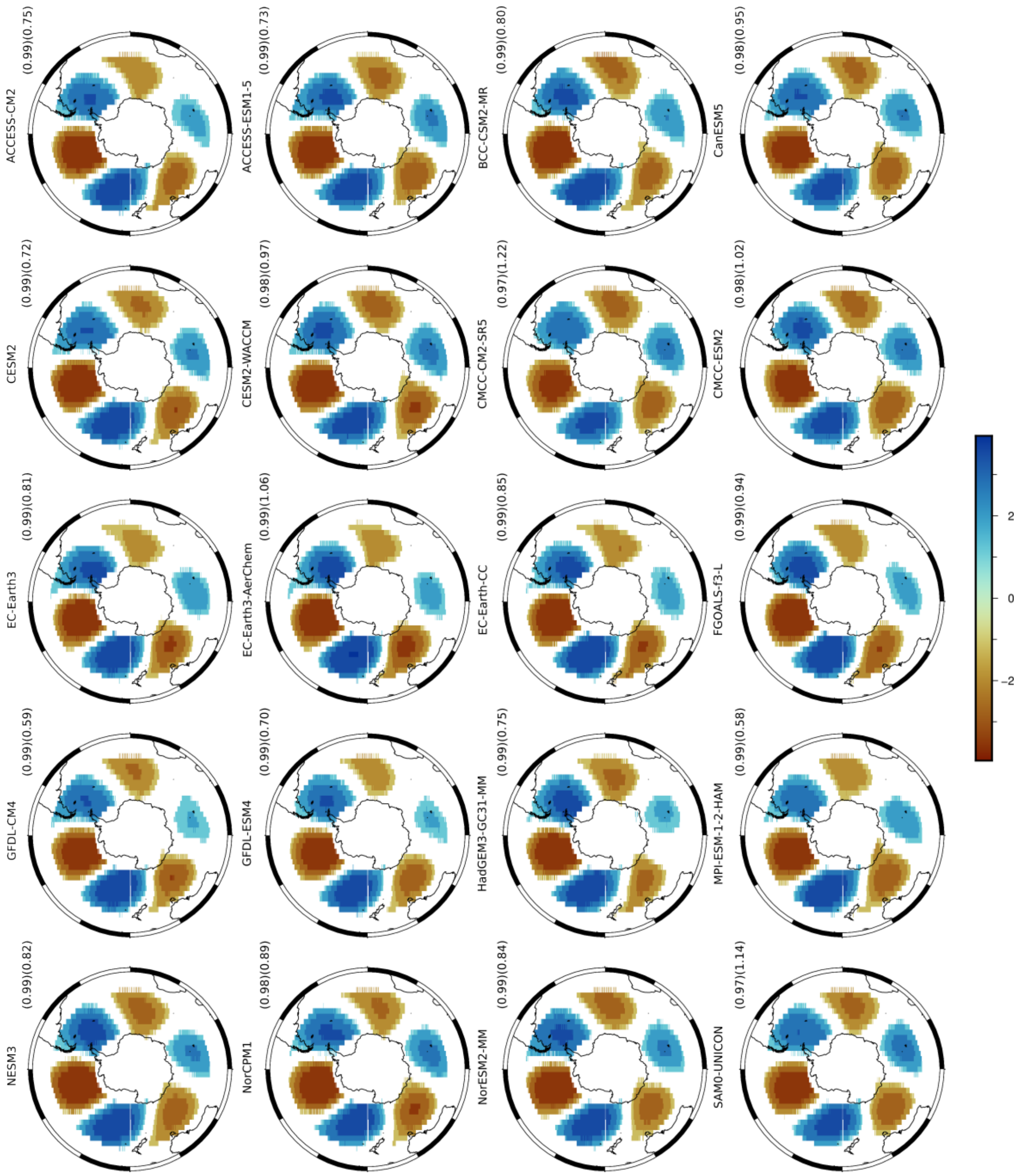

Figure S6. As with Figure S1, but for the Zonal Wave 3, second EOF, values below a magnitude of $|0.5ms^{-1}|$ are supressed.

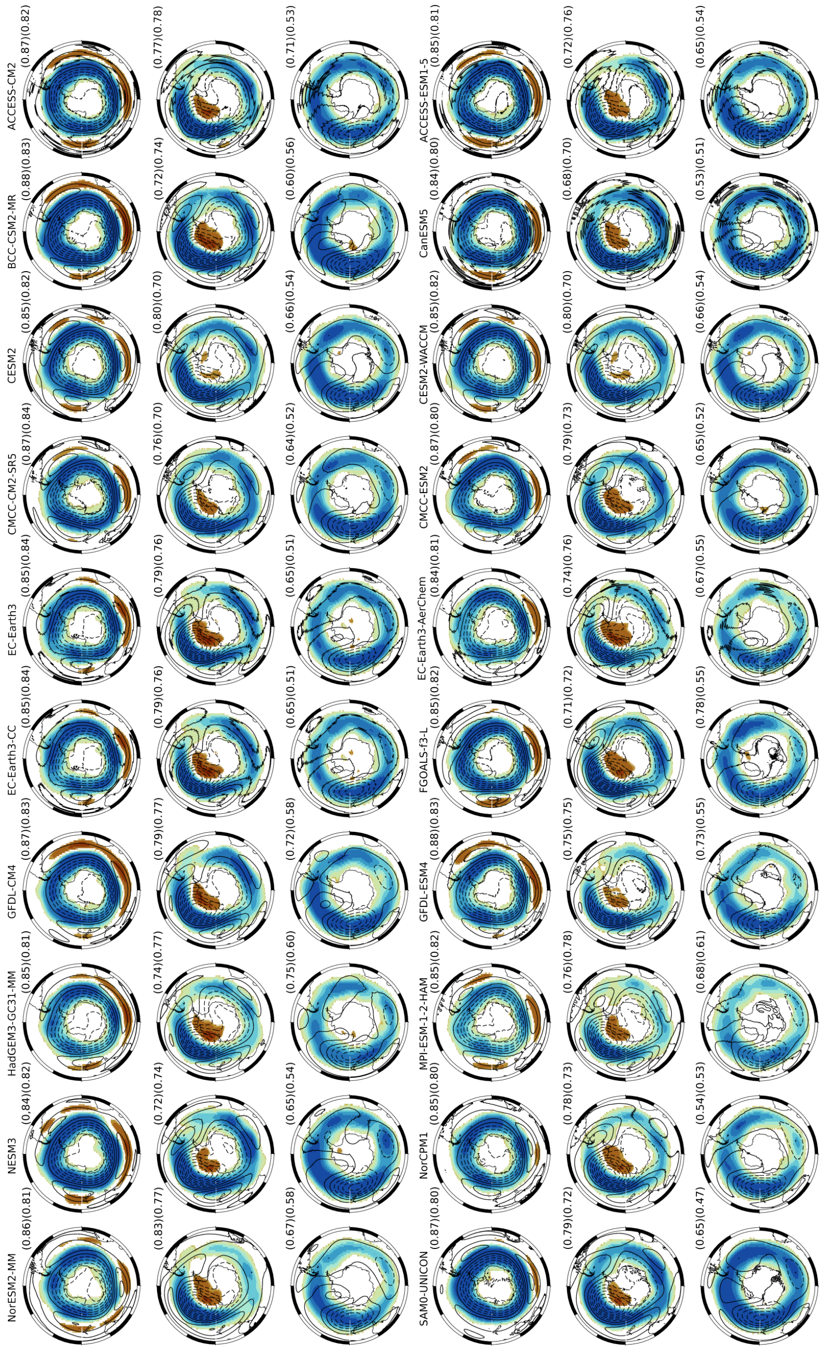

Figure S7. Leading three CBFs of mean Z500 and HF variance fields for each ensemble member for annual data, generated from rotated ERA-5 singular vectors. The mean Z500 field is indicated by the contours ($10m$ intervals), positive contours are solid and negative dashed. HF variance is shown as the colour fill, blue indicating increased HF variance and brown decreased variance. HF variance below $|150m^2|$ is masked.

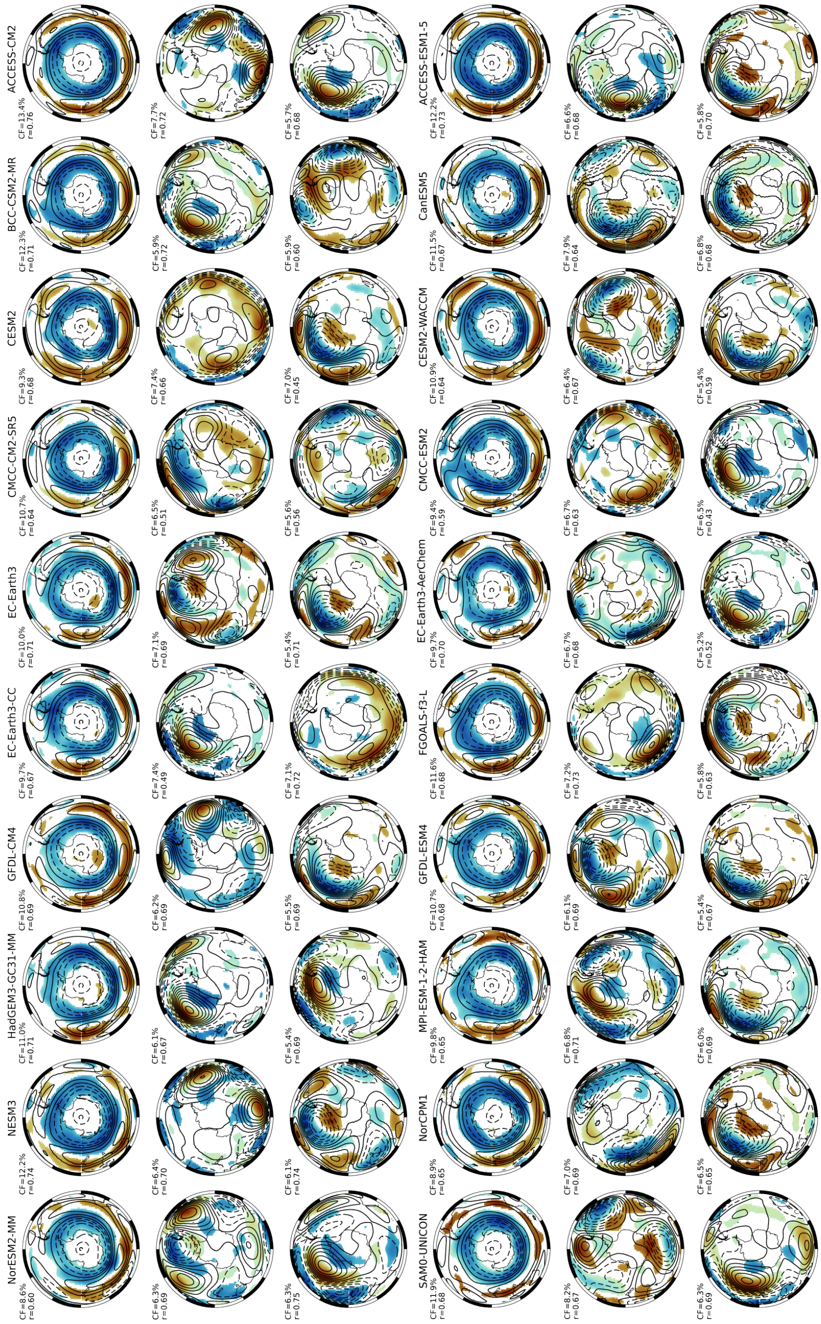

Figure S8. Leading three rotated singular vectors of mean Z500 and HF variance fields from MCA on a monthly timescale, applied to each CMIP6 ensemble member. The mean Z500 field anomalies are indicated by the contours, positive contours are solid and negative dashed. The associated storm track anomalies are shown as the colour fill, blue indicating increased HF variance and brown decreased variance. Units are dimensionless so intervals indicate relative magnitude only.