Development of sea ice diagnostic tools for high-resolution simulations of the Climate Model Intercomparison Project (HiResMIP)

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Cameron Carver
CRVCAM002

Department of Oceanography
University of Cape Town
Rondebosch, Cape Town
South Africa

Supervisors:
Prof Marcello Vichi • University of Cape Town
Dr Dorotea Iovino • Euro-Mediterranean Center on Climate Change

June 2025

Dissertation submitted in partial fulfilment of the requirements for the degree of Master of Science (by coursework and dissertation) in Applied Ocean Sciences, through the Marine and Antarctic Research centre for Innovation and Sustainability (MARiS), in the Department of Oceanography, at the University of Cape Town.

Plagiarism Declaration

I, Cameron Carver, hereby:

1. grant the University of Cape Town free license to reproduce the above thesis in whole or in part, for the purpose of research;

2. declare that:

   (a) this dissertation is my own unaided work, both in concept and execution, and apart from the normal guidance from my supervisor, I have received no assistance except as acknowledged.
   (b) neither the substance nor any part of the above dissertation has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

Signed by candidate

Cameron River Carver

Student's Full Name

Department of Oceanography

University of Cape Town

Wednesday 25 June, 2025

Abstract

Derived from the necessity to more thoroughly understand the role that horizontal resolution plays in the performance of climate modelling, this minor-dissertation describes the development and initial testing of a High-Resolution Sea Ice Diagnostics Toolset. This is designed to evaluate the influence increased horizontal resolution has on the ability of high-resolution climate models to recreate Antarctic sea ice behaviour. Analysis is conducted through the calculation and visualisation of a defined univariate performance metric, referencing a single user defined satellite-derived observational dataset. Developed in Python, cloud-ready datasets and computing capabilities are utilized to eliminate the need for local download of large model datasets and extensive computational capacity. Temporal mean metric values are produced and visualized, providing easy visual analysis of regional model performance for each respective temporal grouping. Spatial mean metric values for each month produce a time series that reveal performance trends. Distributions of these values give insight into the model performance for particular seasons and months. Sensitivity analysis functionality enables the assessment of model performance for specified ranges of observed sea ice concentration values. A preliminary assessment is conducted to substantiate the value of the toolset products. Results indicate alignment with findings of Selivanova *et al*. (2024) where only marginal improvement in model performance is seen with increased resolution for the models assessed. Additionally, sensitivity analysis results highlight the shortcomings of all assessed models in recreating Antarctic sea ice behaviour across the marginal ice zone. Integration of cloud-ready datasets and cloud computing sees significant reduction in time and increased functionality. While the scope of this minor dissertation limits the depth of analysis, the toolset developed here is shown to provide useful insights into the role of horizontal resolution and establishes backbone from which users can generate and visualise univariate metrics for high resolution datasets with minimal local computational load and significantly reduced runtime.

Keywords:

Sea Ice Modelling, High Resolution Datasets, Climate Modelling, Cloud-Ready Data, Remote Repositories, HighResMIP

Acknowledgements

I'd like to thank Prof Marcello Vichi for his guidance, enthusiasm and willingness to introduce and guide an engineer through the intricacies of climate models. I have learned a lot and discovered a whole world of things of which I did not know existed previously.

Thank you to my co-supervisor, Dr Dorotea Iovino, for providing a unique perspective contributing to more extensive feature development.

A special thank you to my family for the never-ending support and for encouraging me to pursue my interests, even if they happen to be halfway around the world.

Table of Contents

List of Figures

List of Tables

List of Symbols/Abbreviations

CMIP – Climate Model Intercomparison Project

ESGF – Earth System Grid Federation

GCM – Global Circulation Model

HighResMIP – High Resolution Model Intercomparison Project

HR – High Resolution

MARiS - Marine and Antarctic Research centre for Innovation and Sustainability

MEF – Model Efficiency Factor

MIZ – Marginal Ice Zone

NOAA - National Oceanic and Atmospheric Administration

NSIDC – National Snow and Ice Data Center

SIC – Sea ice concentration

WCRP - World Climate Research Programme

Chapter 1
Introduction

1.1 Antarctic sea ice in the climate system

Antarctic sea ice plays a vital role in the world climate system, influencing the thermodynamics of climate systems throughout the southern ocean and the rest of the world. A major component of this influence is the result of the albedo effect (Budyko, 1969), the act of a reflective surface, sea ice in this case, returning sunlight to space effectively reducing the amount of heat absorbed inside the atmosphere. Absence of Antarctic sea ice results in exposed ocean, a darker body with comparatively much lower albedo. This open ocean absorbs more solar radiation, leading to increased ocean temperatures modifying the seasonal fluctuation of sea ice expansion and contraction. When this perpetuates as less sea ice, in a given season, more heat is absorbed contributing to a feedback loop (Goosse *et al*., 2018) . The heat flux of the ocean influences ocean circulation and atmospheric dynamics that impact the climate worldwide (Hall, 2004). Acknowledging the influence sea ice has on world climate systems, the ability to predict and understand future impact is a crucial component for the determination of world climate evolution.

1.1.1 Sea ice in climate modelling

Global Circulation Models (GCMs) are extensively complex mathematical simulations of multiple climate systems in an attempt to accurately recreate past and predict future climate behaviour. These models produce simulations of the interaction between climate systems through established thermodynamic, fluid mechanic, biogeochemical and energy transfer equations as applicable (Fig 1-1; Global Circulation Model). By addressing these different properties of all relevant systems in a given area, such as ocean, atmosphere, and sea ice for Antarctic regions, values of climatic characteristics are generated. Resolving these interactions across all areas of the world results in the output of a GCM (McSweeney & Robert, 2018).

Multiple models have been developed by institutions across the world to resolve the dynamic and thermodynamic processes on land, ocean and atmosphere and their coupling. Sea ice is one of the main components simulated by these models. Within models it is defined by key characteristics such as concentration, extent, edge location, thickness, snow depth, and movement vectors. These are the variables that are integrated into GCMs to model the physical and thermodynamic processes. The thermodynamic processes include the freezing and melting of sea ice and energy balance of solar radiation, heat flux between the ocean and the atmosphere and long wave radiation. Physical processes include the drift of the sea ice and how it physically evolves as it grows, shrinks, and breaks apart. All of this is integrated in different ways by each model in an attempt to most accurately recreate and predict the impact that Antarctic sea ice has on the global climate

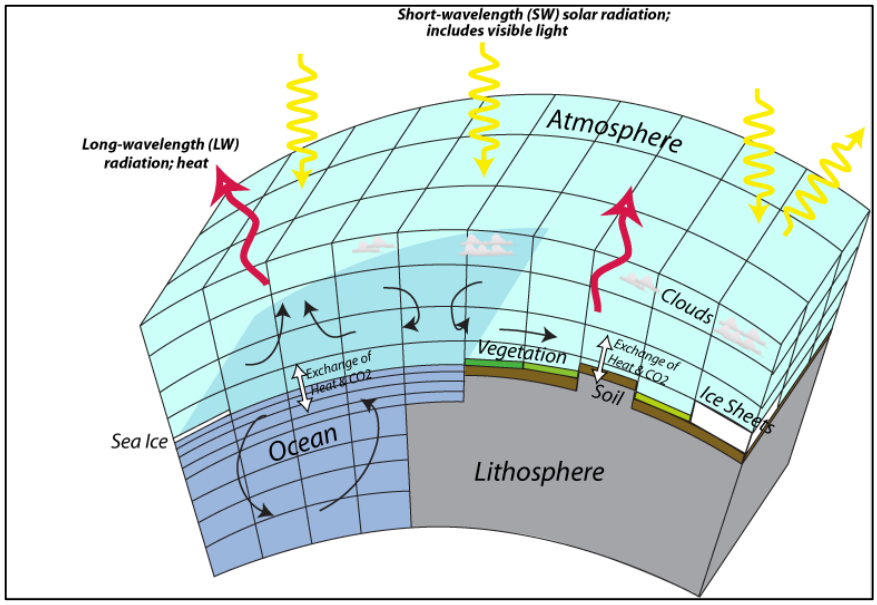


Figure 1-1 The various systems and processes simulated by Global Circulation Models. https://www.appsolutelydigital.com/ModelPrimer/chapter11_section2.html

1.1.2 Model intercomparison projects

Given the complex nature of coupled GCMs, the skill assessment of the various models developed presents a considerable challenge. With the need to establish a level of inter-comparability, the World Climate Research Programme (WCRP) created the Climate Model Intercomparison Project (CMIP; Meehl, 1995) in 1995. This project establishes standards across all models in their nomenclature, output structure and setup to enable inter-comparability. The participating research institutions run systematic concerted experiments of one or multiple main model components, such as atmospheric or oceanic conditions, where different types of boundary conditions are prescribed and the remaining components are allowed to be simulated fully. These forcings provide various insights into potential biases a model may have towards a given component. Various forcing iterations are often run on each model by their respective institutions. The resolution of these different model components has naturally evolved alongside the models themselves with continued technological improvements seen across CMIP1-6 (Meehl, 1995; Meehl *et al*., 1997; Meehl *et al*., 2007; Meehl *et al*., 2009; Meehl *et al*., 2014). Horizontal resolution dictates a model's ability to replicate and predict systems that function on smaller scales than cannot be captured by lower resolution models. With standardization of procedures and products for each model variation enables measurable comparison not only with other models but also with alternate forcing iterations of the same model. This provides a platform from which a deeper understanding of the modelled response to both climate states and variability can be derived. Patterns within and between models can be invaluable insights as to which internal approaches are determining the performance of a given model.

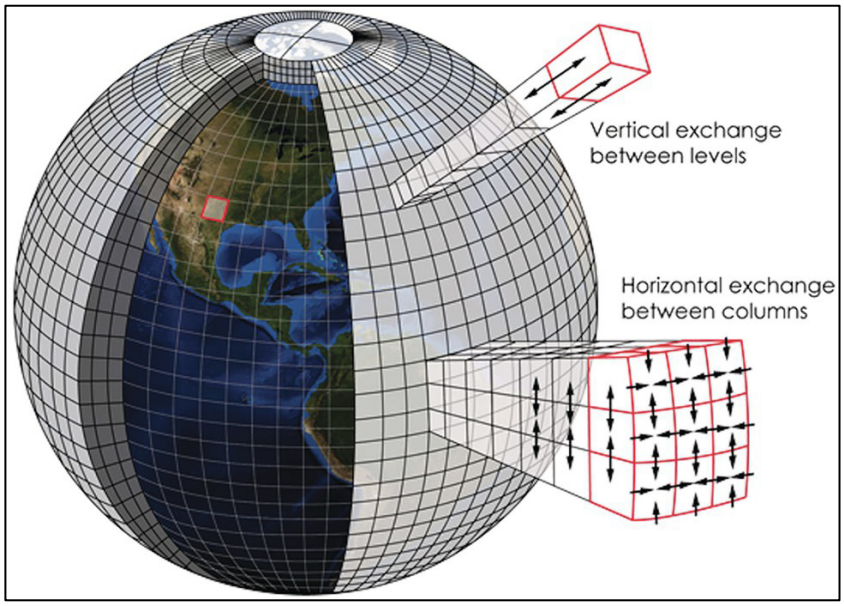


Figure 1-2 Example grid cell structure used by Global Circulation Models, depicting how the earth is divided into discrete sections and the axes in which exchanges are modelled.
https://www.earthmagazine.org/article/todays-weather-forecast-good-strong-chance-improvement

GCMs simulate the climate behaviour by establishing a grid cell structure, like that seen in Fig. 1-2 above, for the entire planet. Each grid cell is treated as a discrete section with uniform values across all inputs and model products. The physical, chemical, and thermodynamic interactions between these discrete grid cells are the core function of GCMs. The size of the established grid structure exponentially impacts the computational capacity required to run a given model. A model which is run based on a 0.25° resolution requires sixteen times the computational capacity of one which is run at a 1° resolution. The complex nature of all the processes modelled by GCMs intuitively lead to the assumption that models ran at a higher resolution will produce more accurate results. This is derived from the notion that increased resolution enables smaller scale systems to be more accurately simulated, where they may have been averaged out with lower resolution. The quantification of this improvement is pertinent as the necessary increased computational capacity directly correlates to an increase in cost and time. Optimising model resolution versus perceived model performance also optimises the use of computational resources. It is valuable to know if a considerable increase in horizontal resolution only results in a negligible increase in model performance as this would indicate that the additional computational resources required for high resolution models are not worthwhile. Increased efficiency in the use of computational resources enables more analysis to be completed over the same time period, improving the rate of development.

With the aim to provide a multi-model approach to the investigation of model performance improvement potentially observed with increased horizontal resolution, the High-Resolution Model Intercomparison Project (HighResMIP) was established. Endorsed by CMIP6, HighResMIP conducts investigative experiments on model simulations at standard

and higher resolutions of atmosphere and ocean components. The experiments include atmospherically forced and coupled runs within the time period of 1950-2050 (Haarsma *et al*., 2016). Results of HighResMIP are intended to provide insight to the potential benefit of increased horizontal resolution this can then be used to consider the validity of additional costs attributed to the corresponding increased time and computational load.

1.1.3 Impact of resolution on Antarctic sea ice simulations

Analysis completed by Selivanova *et al*. (2024) explored the impact of increased horizontal resolution on various models' representations of Antarctic sea ice from HiResMIP and concluded that the benefits are limited. They suggest that the difference in performance observed between model runs of different resolutions is attributed more to the interaction of the atmospheric and ocean components than to the change in horizontal resolution. With no systemic beneficial effect of increased horizontal resolution observed across the HighResMIP models, the intuitive notion of increased horizontal resolution directly improving a model's performance is not substantiated in the case of Antarctic sea ice simulations. Selivanova *et al*. (2024) also reviewed this same potential impact regarding Arctic Sea Ice simulations and again observed limited improvements in model performance with increased horizontal resolution. Though they did note that increased ocean resolution did show more improvement in model performance than increased atmosphere resolution. While this thesis focuses on Antarctic sea ice it is feasible that the developed toolset could be applied to Artic sea ice as needed.

The counterintuitive nature of these findings furthers the necessity to better understand the impact of increased horizontal resolution with multivariate analyses. Drawing from the success of CMIP and HighResMIP the development of a tool capable of high-resolution model intercomparison, through established metrics, can provide useful insight into these irregularities. With the aim to develop a versatile toolset that provides efficient and flexible high resolution model analysis, existing toolsets and techniques for high resolution data handling are explored.

1.2 Objectives

The aim of this paper is to present the diagnostics toolset developed to provide a high-resolution compatible backbone from which any univariate performance metric can be calculated and visualised to provide further insight into the impact of increased horizontal resolution on GCM performance. Reduction of computational load and time is sought through the utilisation of cloud-ready data and remote computing capabilities. The Pangeo remote repository is integrated into the new toolset to achieve a substantial reduction in local storage requirement, computational load and runtime. The univariate metric, Nash-Sutcliffe model efficiency coefficient (Nash *et al*., 1970), is selected for initial demonstration of the toolset's functionality. A walkthrough of the toolset's structure is provided for both the calculation and visualisation of the relevant metric. Initial testing is completed with twenty four CMIP6 models before six HighResMIP models are run through the toolset. Spatial and temporal mean

values are produced as well as an aggregated representative performance value for the cumulative model and each season and month. A sensitivity analysis is also completed to provide insight into model performance relative to specific levels of sea ice concentration. Results are visualised and a preliminary analysis completed to corroborate the insights that can be revealed through the use of this toolset. It is intended that this toolset be utilised to further delve into the influence of increased horizontal resolution on GCM performance.

1.3 Model evaluation

1.3.1 Model performance of sea ice

The most common sea ice state variables or diagnostics produced by models are concentration, extent, edge location, thickness, snow depth, and drift. These provide the information to sufficiently establish the physical and thermodynamic influence that sea ice has on the global climate. Of these characteristics, concentration and extent are most frequently used as large-scale metrics for the evaluation of a model's ability to reproduce and predict sea ice behaviour. Sea ice concentration (SIC) is selected as the test case for this first implementation of this toolset.

All model evaluation methods are based on available observational data. Observational datasets, with sufficient spatial and temporal coverage for Antarctic SIC, are derived from satellite observations, with algorithms applied to raw passive microwave data to derive the physical variables of interest. Model evaluation consists of deriving a metric value from a model's behaviour in comparison to the observational dataset using one or multiple statistical parameter(s) of each depending on the specific metric of interest. The processes used to derive observational datasets have inherent uncertainty and sea ice concentration values at or below 15% are commonly considered to have too high uncertainty to be used as reference (Cavalieri *et al*., 1991). Observational values below this threshold can still be evaluated but the inherent uncertainty in the reference dataset, from which the evaluation metrics are derived, must be acknowledged during analysis.

The Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4 (Meier *et al*., 2021) from the National Oceanic and Atmospheric Administration's (NOAA) National Snow and Ice Data Center (NSIDC) is chosen as the observational reference dataset for this implementation of the toolset. The toolset is designed to work with any observational dataset with monthly frequency and a sufficiently defined spatial grid.

1.3.2 A diagnostic toolbox for SIC: SITool

When exploring existing diagnostics approaches the Sea Ice Evaluation Tool (SITool; Lin *et al*., 2021) was identified as a candidate from which a high-resolution analysis could potentially be produced. The SITool generates metrics for each of the key sea ice characteristics listed in Sec. 1.1.1; concentration, extent, edge location, drift, thickness, and snow depth. The SIC metric is based on scaled absolute errors and utilises two observational datasets; NSIDC-0051 (Cavalieri *et al*., 2021) and OSI-450 (OSI SAF, 2017). The original application of SITool was conducted on nine CMIP6-OMIP models. The computation structure

of SITool is based on the python package NumPy, a useful package for handling datasets in general but one that has limitations when addressing large datasets seen with high-resolution models. NumPy operates linearly and is optimised for calculations to be performed on the local memory of the users device. This becomes infeasible for large datasets that can exceed available RAM capacities. The initial intent was to adapt the SITool for use with high-resolution datasets through the integration of XArray. XArray is a python package designed specifically to handle very large datasets, optimising its calculation structure for parallel computing across multiple cores.

1.4 Cloud-ready climate data and Pangeo

To address the complications of extensive computational capacity and processing time, cloud-ready data and cloud computing capabilities were explored. Cloud-ready data are datasets that have been optimised for cloud storage and computation such that an analysis can be directly conducted in parallel on multiple 'chunks' from a subset of a larger dataset, thus eliminating the need for the analysis to navigate through the entire dataset each time (Stern *et al*., 2022). Utilising cloud computational capacity in conjunction with the cloud-ready data provides many benefits. These include increased efficiency, reliability, reproducibility and reduced costs as outlined by Abernathey *et al*. (2021). Increased efficiency is a result of exploiting the extensive computational power that exists in cloud-based servers, a capacity that vastly outperforms any local device likely to be used by those conducting analyses. Similarly, by not having to locally download any data there is no potential of a bandwidth bottleneck increasing runtime. Cloud servers are well established and maintained by their respective providers, leading to increased reliability and safety of the data, this also lends itself to ensure reproducibility across different assessments. With everyone accessing the same dataset there can be a great reduction in instances of datasets becoming inaccessible in the specific format utilised by an individual on their local device. Ensuring everyone has access to the same datasets also increases the ease of collaborative efforts. All of this coalesces into a far more efficient workflow than what has been traditionally conducted with local download and computation.

This increased efficiency becomes even more pertinent as model resolution increases. While low resolution model datasets are typically multiple hundreds of megabytes, for a single variable at monthly frequency, high resolution model datasets can be multiple gigabytes for the same variable and frequency. When looking to analyse multiple variables across multiple models the data capacity required to store the datasets locally compounds quickly. By utilising a remote repository and cloud computing, calculations can be completed remotely on datasets that have been pre-processed for cloud computation compatibility. This drastically reduces the storage requirement of the user's local device to only that of any reference dataset and any output generated by the analysis. With potential future inclusion of reference datasets on a cloud repository this storage requirement can be reduced to only the resulting analysis output.

One such remote repository is a part of the Pangeo project (https://www.pangeo.io/). Pangeo is a community of scientists and programmers that are actively developing software and platforms to benefit the geoscientific community. Pangeo's remote repository optimises and hosts cloud-ready climatological datasets and facilitates calculations through cloud computing resources.

Though cloud-ready datasets are structurally manipulated to enable remote access and efficient cloud-based assessment, the standardisation of internal nomenclature must also be ensured for functional intercomparison. The xMIP package (Busecke *et al*., 2024) was developed to remedy these inconsistencies. This open-source python package is a preprocessing package for GCM datasets within the Pangeo repository. The xMIP package intakes a dataset and remedies inconsistencies in naming conventions, location of grid coordinate information and units. Through xMIP all model datasets output can confidently be uniformly compared. Through the integration of Pangeo, and preprocessing with xMIP, cloud ready datasets are utilised in this new toolset and the anticipated increased efficiency becomes evident.

Chapter 2
Methods

2.1 SITool adaptation

The SITool (Lin *et al*., 2021) was not originally developed to handle the additional computational load that high resolution data demands. This was remedied with the integration of the XArray package. SITool was downloaded from the GitHub repository, https://github.com/XiaLinUCL/Sea-Ice-Evaluation-Tool, and the sea ice concentration diagnostics was firstly used with the suggested test cases from the Ocean Model Intercomparison Project (OMIP; Griffies *et al*., 2016) model outputs. Testing was then conducted with historical CMIP6 models all of which resulted in extensive computational runtimes and often in the crashing of the computer kernel. The initial integration of XArray into SITool was successful at enabling the import and running of preliminary calculations. Data from individual historical CMPI6 and OMIP models were downloaded from the ESGF repository (Cinquini *et al.,* 2014) utilising the following search criteria:

Select a Project – CMIP6 or OMIP
Filter by Transfer Options – Only Globus Transferrable
General - Activity ID – CMIP
Identifiers – Experiment ID – historical
Labels – Grid Label – gn
Classifications – Table ID – Simon
Classifications – Variable ID – SIConc.

Further integration of XArray was abandoned when reviewing the basis of the metric produced by the SITool. As mentioned in Sec. 1.3.2, the metric calculated by the SITool to assess model performance of SIC is based on the absolute difference in errors between two observational datasets. The resulting metric values are a scaled ratio between two absolute differences. The first being the absolute difference in error between the model prediction and observational datasets, and the second being the 'typical error' defined by the absolute difference between two observational datasets (Lin *et al*., 2021). The metric is scaled off of how well the model's uncertainty relative to an observational dataset compares to the uncertainty between two observational datasets. Utilisation of this metric long-term is likely not feasible as the difference in the measure of uncertainty between observational datasets tends to converge with continued refinement of algorithms and technologies from which they are derived. This has been observed in an analysis of SIC datasets regarding the marginal ice zone (Vichi, 2022).

2.1.1 Alternative model evaluation metrics

Multiple model evaluation metrics exist; ones derived from a single observational dataset were identified to eliminate the future usability issues seen with SITool. The metric used in the example configuration, explained below, is the Nash-Sutcliffe model efficiency coefficient (Nash *et al*., 1970), also known as the Model Efficiency Factor (MEF). This metric provides a scale for how a model performs in reference to an observational dataset, establishing how well the model performs relative to mean observational values. Multiple other metrics exist including Brier score (Brier, 1950), Kling-Gupta Efficiency (Gupta *et al*., 2009), and Murphy's skill score (Murphy, 1973). The toolset is developed so that it can be adjusted to calculate and plot any univariate model performance metric as desired.

2.2 New tool development: technical description

To overcome the issues highlighted in Sec. 1.3-1.4 and in the previous section, a new toolset is developed from the ground up to produce analytical results from a defined univariate model performance metric through the use of cloud-based datasets and computational capacity. The toolset consists of two python notebook types, "metric calculation" and "metric plotting", these have two and four variations respectively for different applications explained in Sec. 3.1.1. The toolset was developed in python 3.12.7 and utilises the following packages: `os`, `copy`, `time`, `NumPy`, `intake`, `matplotlib`, `XArray`, `xESMF`, `clisops`, `cartopy`, and `xMIP`.

The calculation notebooks produce the calculated metric as temporal annual, seasonal, and monthly means as well as a spatial mean time series. The plotting notebooks intake the calculated metric values and plot the temporal mean annual (Figs. 3-2, 3-7), seasonal (Figs. 3-3, 3-8), and monthly (Figs. 3-4, 3-9) values on a southern polar stereographic projection with appropriate performance categorisation colour mapped accordingly. The monthly spatial mean values are plotted as a time series (Figs. 3-5, 3-10) as well as being binned into seasonal (Figs. 3-6(a-b), 3-11(a-c)) and monthly (Figs. 3-6(c-d), 3-11(d-f)) categories and plotted as stacked bar graphs. The aggregated values are calculated by taking the temporal mean of the spatial mean monthly values and plotting them as heat mapped blocks (Fig. 3-12) with each value for annual, seasonal and monthly displayed in their respective boxes.

The filesystem structure for both notebook types is outlined in Fig. 2-1 below. The observational reference dataset and its accompanying reference grid dataset are stored under the `Observations` directory. Outputs from the calculation scripts are directed to the appropriate `CMIP-MEF` or `HRMIP-MEF` directory. All plots are exported to appropriate directories within the `Plots` directory. The `Combined` directory stores all plots generated from the `plot_cumulative.ipynb` script later described in Sec 3.1.1. Spatial and temporal mean plots of values generated by equations defined in Sec. 2.2.6 and Sec. 2.2.7 are saved to the `Spatial` and `Temporal` directories respectively. All plots resulting from the

sensitivity analysis configurations defined in Sec. 2.2.9 are stored to the `Sensitivity` directory.

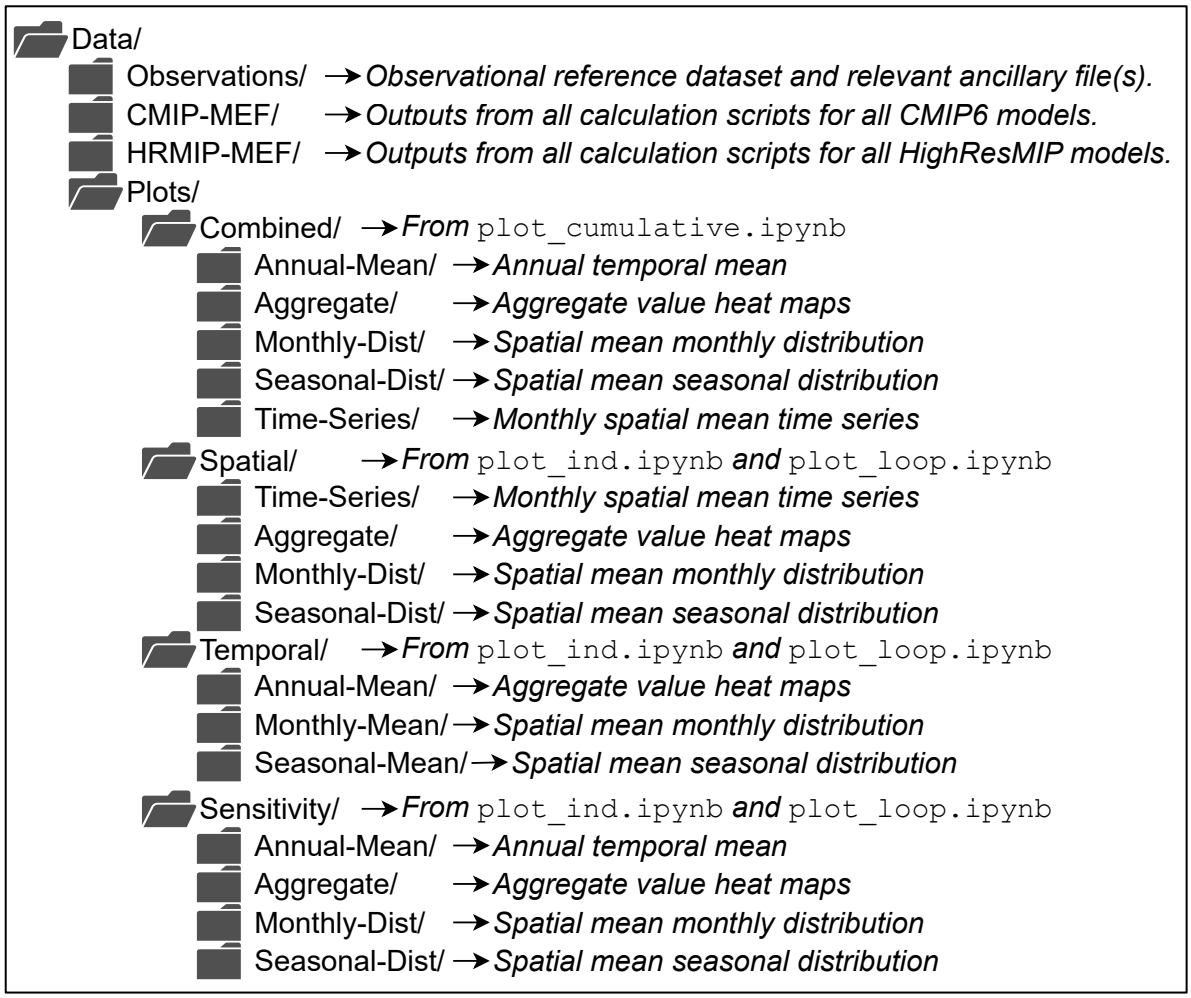


Figure 2-1 File system structure of the new toolset indicating location of all inputs and outputs.

The overview of workflow conducted by the calculation script is outlined below in Fig. 2-2 and each component explained in Sec. 2.2.1 through Sec 2.2.9.

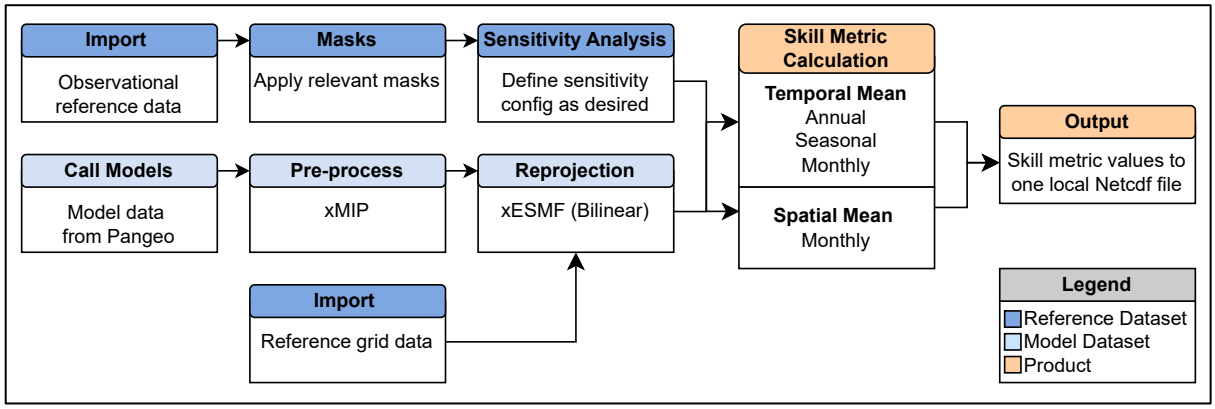


Figure 2-2 Flowchart of the processes conducted by the metric calculation notebook.

2.2.1 Import - Reference data

The choice of reference data depends on the preference of the user. The toolset allows for the use of any climatological observational data with sufficiently defined grid data and a temporal frequency that matches that of the model(s) of interest. For this first iteration we selected the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4 (Meier *et al*., 2021). This was downloaded from the NSDIC repository at (https://nsidc.org/data/g02202/versions/4). This dataset is also used in SITool, and is composed of monthly sea ice concentration data provided as individual monthly files for the period October 1978 through March 2024. These are combined, along the temporal axis, using the `netCDF-combiner.py` script available at the GitHub repository: https://github.com/CCarverUCT/HighRes-Sea-Ice-Diagnostic-Toolset. The resulting file is named `seaice_conc_monthly_sh_n07_v04r00_all.nc`. The corresponding reference grid file, `G02202-cdr-ancillary-sh.nc`, is also downloaded from the NSIDC repository (https://noaadata.apps.nsidc.org/NOAA/G02202_V4/ancillary/). These are the only data that need to be downloaded to the user's local device.

The calculation script imports the reference data and corresponding ancillary file, applying a land-sea mask for all ocean values based on the ancillary coordinate grid file. This reference data is referred to as the observational data moving forward. The observational data is loaded into the workspace and a time slice taken for the period of December 1978 through November of 2014 for a total of 432 months. This takes full advantage of the historical CMIP6 model data, which currently runs up until 2015. The selection starts in December and concludes in November to align with the seasonal categorisation of a December start. Defined flagged values within the data are masked out. A count of the quantity of values greater than zero at each grid point is conducted. Each grid point that contains less than or equal to twenty values matching this criterion, over the course of all 432 months, is masked out. This eliminates grid points that have recorded observed sea ice presence less than 5% of the time, a behaviour that is not intended for a model to simulate. By removing these occurrences, a model is not penalised for inaccurately modelling these outlying grid points. A temporal mean of the observational SIC data is taken.

2.2.2 Call models – Pangeo repository

Models of interest are listed to streamline the inquiry into the remote repository. Both CMIP6 and HighResMIP model lists are generated. CMIP6 models accessed are provided in Table 1 below. HighResMIP models accessed and their respective nominal resolutions for ocean and atmosphere components are provided in Table 2 below. These six HighResMIP models were selected for this initial proof of concept as they were the first models available on the Pangeo repository and provide sufficient evidence of useful insights derived from the new toolset.

The Pangeo catalogue address is defined. There are multiple catalogues available, and the user must uncomment, or self-define, the one that is suitable for their analysis. During development no HighResMIP models were initially available on the Pangeo repository. A

request was submitted to the CMIP6-Leap-Feedstock (https://github.com/leap-stc/cmip6-leap-feedstock/issues/180) and multiple HighResMIP models were added to the Pangeo repository within one month. With the catalogue defined the search parameters are further specified and running this produces a catalogue of all available matching datasets.

CMIP6 Models	Data Reference
ACCESS-CM2	Dix, M *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.4271.
ACCESS-ESM1-5	Ziehn, T *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.4232.
BCC-CSM2-MR	Wu, T *et al*. (2018). http://doi.org/10.22033/ESGF/CMIP6.2948.
BCC-ESM1	Zhang, J *et al*. (2018). http://doi.org/10.22033/ESGF/CMIP6.2949.
CAMS-CSM1-0	Rong, X (2019). http://doi.org/10.22033/ESGF/CMIP6.9754.
CanESM5	Swart, N.C *et al*.(2019). https://doi.org/10.22033/ESGF/CMIP6.3610.
CAS-ESM2-0	Chai, Z (2020). https://doi.org/10.22033/ESGF/CMIP6.3353.
CMCC-CM2-SR5	Lovato, T *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.3825.
CMCC-ESM2	Lovato, T *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.13195.
FGOALS-f3-L	YU, Y (2019). https://doi.org/10.22033/ESGF/CMIP6.3355.
FGOALS-g3	Li, L (2019). https://doi.org/10.22033/ESGF/CMIP6.3356.
FIO-ESM-2-0	Song, Z *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.9199.
GFDL-CM4	Guo, H *et al*. (2018). http://doi.org/10.22033/ESGF/CMIP6.8594.
GFDL-ESM4	Krasting, J P *et al*. (2018). https://doi.org/10.22033/ESGF/CMIP6.8597.
MIROC6	Tatebe, H, Watanabe, M (2018). https://doi.org/10.22033/ESGF/CMIP6.5603.
MPI-ESM-1-2-HAM	Neubauer, D *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.5016.
MPI-ESM1-2-LR	Wieners, K-H *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.6595.
MRI-ESM2-0	Yukimoto, S *et al*. (2019). http://doi.org/10.22033/ESGF/CMIP6.6842.
NESM3	Cao, J, Wang, B (2019). http://doi.org/10.22033/ESGF/CMIP6.8769.
NorCPM1	Bethke, I *et al*. (2019). http://doi.org/10.22033/ESGF/CMIP6.10894.
NorESM2-LM	Seland, Ø *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.8036.
NorESM2-MM	Bentsen, M *et al*. (2019). https://doi.org/10.22033/ESGF/CMIP6.8040.
SAM0-UNICON	Park, S, Shin, J (2019). https://doi.org/10.22033/ESGF/CMIP6.7789.
TaiESM1	Lee, W-L and Liang, H-C (2020). https://doi.org/10.22033/ESGF/CMIP6.9755.

Table 1 CMIP6 models used for initial testing and their corresponding references.

HighResMIP Model	Config	Resolution		Data Reference
		Ocean	Atm.	
CMCC-CM2	HR	0.25°	100km	Scoccimarro, E *et al*. (2019). http://doi.org/10.22033/ESGF/CMIP6.3817.
	VHR	0.25°	25km	Scoccimarro, E *et al*. (2018). http://doi.org/10.22033/ESGF/CMIP6.3818.
NOAA-GFDL	CM4C192	0.25°	25km	Zhao, M *et al*. (2018). http://doi.org/10.22033/ESGF/CMIP6.8567.
HadGEM3	LL	1°	250km	Roberts, M (2017). http://doi.org/10.22033/ESGF/CMIP6.6042.
	MM	0.25°	100km	Roberts, M (2017). http://doi.org/10.22033/ESGF/CMIP6.6045.
	HM	0.25°	50km	Roberts, M (2018). http://doi.org/10.22033/ESGF/CMIP6.6040.

Table 2 HighResMIP models assessed by the toolset and their corresponding references.
Also their respective nominal atmosphere and ocean resolutions defined. Modified from Selivanova *et al*. (2024).

2.2.3 Pre-processing - xMIP

All models in the catalogue are called and pre-processed through the xMIP package (Busecke *et al.*, 2024) to normalise and ensure proper internal nomenclature. The list of pre-processed models is then displayed for final confirmation prior to reprojection and respective model names extracted for future reference.

When calculating MEF values for a single model the model native grid and the variable `siconc` are plotted side by side to visually confirm logical data. This is not done when addressing multiple models to reduce overall computation load and runtime.

2.2.4 Reprojection - xESMF

The ancillary NSIDC file contains the relevant grid information for the NSIDC observational data and is used as the standard grid for MEF metric spatial calculations. The coordinate reference system defined is WGS 84 / NSIDC Sea Ice Polar Stereographic South as specified by ESPG:3976 (https://epsg.io/3976). This is treated as the target grid and all high-resolution model data is reprojected onto this.

With all models pre-processed, each dataset is time sliced to match that of the observational dataset. Relevant values are masked out as specified in the metadata and extraneous dimensions are excluded to reduce dataset size. Model names are extracted from the defined source id.

Utilising the xESMF package (Zhuang *et al.*, 2024) the reference grid is defined as the target grid and a bilinear reprojection is completed for each model's data. The land mask defined in the reference grid file is overlayed to ensure alignment with the observational data.

Time coordinates are aligned as many model data are month-start aligned while the observational reference is month-middle aligned.

For calculation of a single model's MEF values the `siconc` variable of both the model and reference dataset are plotted for a single month to visually confirm successful reprojection and proper scaling of `siconc` variable to be within a range of 0-1, representing 0-100%.

2.2.5 Model efficiency performance index

The model efficiency performance index (MEF) proposed by Nash and Sutcliffe (1970) is selected as the main metric for the initial iteration of this toolset. This index measures the ratio of the model data error to the variability of the observational data (Eq. 1). The resulting values are less than or equal to 1 with value-based performance categorised as: (0.65,1] 'excellent', (0.5,0.65] 'very good', (0.2,0.5] 'good', and (0,0.2] 'poor'. Any MEF value less than 0 represents an instance where the model performs worse than the observational mean and are categorised as 'bad'. A MEF value equal to one represents a model that perfectly represents the observational data.

Any metric calculation based off a single observational reference can be substituted in place of the MEF equation and the tool will produce the appropriate results. An open-source library named SkillMetrics (Rochford, 2016), that contains multiple skill indices, was also tested, but it was not used further as it did not allow for seamless integration due to not being structured for compatibility with multidimensional data. Any integration of this metric library would require major changes.

2.2.6 Temporal mean MEF calculation

Cumulative temporal mean MEF values are calculated at each grid point. These are calculated with the equation below (Eq. 1), where O_n and P_n are the observational and modelled values at each discrete grid point for each month n across N months.

$$MEF = 1 - \frac{\sum_{n=1}^{N}(O_n - P_n)^2}{\sum_{n=1}^{N}(O_n - \bar{O})^2} \quad (1)$$

Seasonal temporal mean MEF values are calculated at each grid point by grouping the numerator and denominator values into DJF, MAM, JJA, SON seasonal groupings, s, before conducting the summation over S months. This equation is below:

$$MEF_s = 1 - \frac{\sum_{s=1}^{S}((O_s - P_s)^2)_{grouped\ by\ season}}{\sum_{s=1}^{S}((O_s - \bar{O})^2)_{grouped\ by\ season}} \quad (2)$$

Monthly climatological MEF values are calculated at each grid point by grouping the numerator and denominator values into the twelve-month groupings, m, before conducting the summation over, M, months. The equation is below:

$$MEF_m = 1 - \frac{\sum_{m=1}^{M}((O_m - P_m)^2)_{grouped\ by\ month}}{\sum_{m=1}^{M}((O_m - \bar{O})^2)_{grouped\ by\ month}} \quad (3)$$

Both the seasonal and monthly temporal mean MEF values are calculated using the overall observational mean, $\bar{O}$, such that the resulting values reflect their respective contribution to the cumulative mean MEF values.

2.2.7 Spatial mean MEF calculation

Spatial mean MEF values are calculated for each month by taking the summation of the numerator and denominator across all grid points for each individual month across the entire time period defined. This provides a single representative MEF value, of all grid points with valid observational values, for each temporal datapoint. This is used to create a time series plot to visualise potential trends in seasonal, monthly or long-term model performance. These values are also binned into seasonal and monthly groups and MEF performance distribution for each respective group can be analysed. The equation for this calculation is below, where x and y represent all grid points on the defined reference grid with size X,Y.

$$MEF_b = 1 - \frac{\sum_{y=1}^{Y}\sum_{x=1}^{X}\left((O_{xy} - P_{xy})^2\right)_{xy}}{\sum_{y=1}^{Y}\sum_{x=1}^{X}\left((O_{xy} - \bar{O})^2\right)_{xy}} \quad (4)$$

2.2.8 Data export

All calculated MEF metrics are saved to single Netcdf file with the naming structure of `{model}-MEF.nc`. These are saved to the appropriate directory outlined in Fig 2-1. The variable nomenclature and resulting dimensions for each model are defined in Table 3 below.

Variable	Note	Dimensions
MEF_a	Annual mean spatial MEF values	{y: 332, x: 316}
MEF_s	Seasonal mean spatial MEF values	{season: 4, y: 332, x: 316}
MEF_m	Monthly mean spatial MEF values	{month: 12, y: 332, x: 316}
MEF_b	Temporal MEF values	{tdim: 432}
model	Model name	-

Table 3 Variables exported by the MEF calculation notebook and their corresponding dimensions for the initial test run.

2.2.9 Sensitivity analysis

Sensitivity analysis of how a model reproduces sea ice in specific concentration ranges are run to further investigate the performance of each model depending on the observed SIC state. The default configuration generates metric values based on all observed sea ice concentration values [0,100]. The different sensitivity analyses are created by limiting the defined observational dataset to values of specific ranges before proceeding with the metric calculations. The results produced by each analysis are then representative of the models performance for the defined range of observed values. To retain comparability across the metrics produced by these different analyses the observational mean, $\bar{O}$, remains the same and is calculated as the mean for all observations in the default range of [0,100]. The default configuration and five specific sensitivity analysis observational ranges are defined in Table 4 below. The range of (15,100] isolates all observational SIC values that are above the 15% SIC uncertainty threshold defined in Sec. 1.3.1. The behaviour of a model for values under this threshold is ascertained with the (0,15] observational SIC range. Isolating (0,100] assesses the models ability to predict only when there is sea ice present, disregarding open ocean. The (15,80] observational SIC range represents that of the marginal ice zone (MIZ) and (80,100] is considered to be pack ice.

Naming	Observational Range	Naming	Observational Range
DEF	[0,100]	I015	(0,15]
I15100	(15,100]	I1580	(15,80]
I0100	(0,100]	I80100	(80,100]

Table 4 Sensitivity analysis variation naming and observation sea ice concentration value range.

2.3 MEF plotting

The visualisation of metrics calculated with the calculation scripts defined in Sec. 2.2 is produced with the plotting scripts as outlined below in Fig. 2-3. The specifics of this workflow are described in Secs. 2.3.1 – 2.3.4 below.

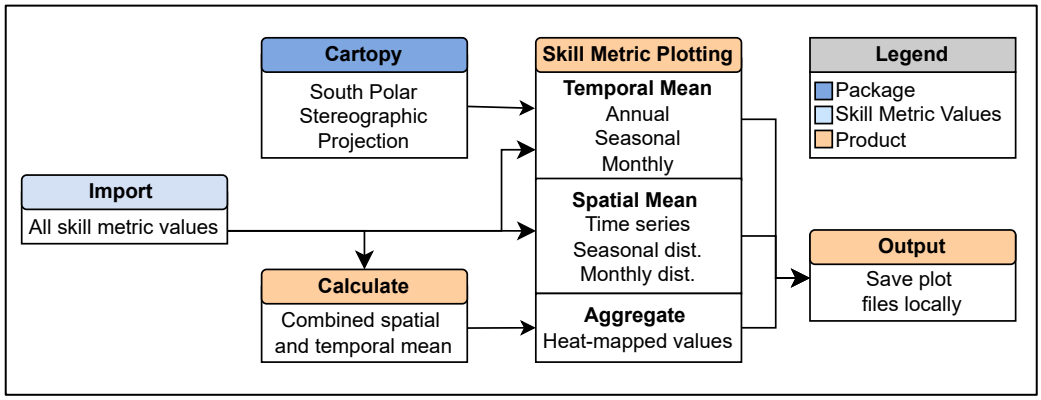


Figure 2-3 Flowchart of the processes conducted by the metric plotting notebook.

2.3.1 Data import

Files containing the relevant metric values are all loaded into the workspace from the user's local device. This procedure is repeated in both files to ensure that they can be executed separately. Relevant variables and metric values are extracted from the imported dataset. All relevant packages are imported and initial colour-mapping levels and colours declared.

2.3.2 Temporal mean plots

Annual, seasonal and monthly MEF temporal mean values are plotted, with the Cartopy package, on a southern polar stereographic projection using the categorisation of the metric values specified in Sec. 2.2.5 depicted as unique colours. Coastlines, latitudes and longitudes are included in all plots. The model's name is automatically included in all plots, drawing from the stored model name from the original metadata.

2.3.3 Spatial mean plots

The monthly spatial mean metric values are plotted as a time series plot with the metric categories, specified in Sec. 2.2.5, colour coded and underlaid. For this configuration displaying the MEF metric, the metric axis is restricted to [-2, 1] to enable direct comparison of all models regardless of individual range in MEF values.

These time series metric values are then binned into seasonal and monthly categories and plotted on a stacked bar graph, with the appropriate colour coding applied corresponding to categorical grouping defined in Sec. 2.2.5. The total count for each season or month should equal the number of years spanned by the data used to calculate the metric.

2.3.4 Aggregated metric values

Aggregated metric values are calculated in the plotting notebook to eliminate redundant calling of model data as these values are based off the previously calculated metric of monthly spatial mean values `MEF_B`. The mean of the spatial mean values is taken across the relevant temporal groupings of annual (*N*), seasonal (*S*), and monthly (*M*). The equations used to calculate these values are below:

$$MEF_{Annual} = \frac{1}{N}\sum_{n=1}^{N} MEF_b \tag{5}$$

$$MEF_{Season} = \frac{1}{S}\sum_{S=1}^{S} (MEF_b)_{grouped\ by\ season} \tag{6}$$

$$MEF_{Month} = \frac{1}{M}\sum_{m=1}^{M} (MEF_b)_{grouped\ by\ month} \tag{7}$$

These three equations each produce a single aggregated performance index value, for the respective group of interest; annual (Eq. 5), each season (Eq. 6) and each month (Eq. 7). This is calculated from both temporal and spatial averaging. These are visualised on a colour coded heat map.

All plots generated can be saved to a user defined directory with built in naming structure calling from the model name drawn from the imported metric values file.

Chapter 3
Results

3.1 Technical features of the new toolset

3.1.1 Jupyter notebooks

Two variations of the calculation script described in Sec. 2.2 were developed: a `Calc_Ind.ipynb` script to address a single model and a `Calc_Cumulative.ipynb` script to sequentially address each model in a defined list and output a single Netcdf file.

Four alternate plotting notebooks are used for different applications as described in the below Table 5.

`Plotter_Ind.ipynb`	Plots output of `Calc_Ind.ipynb` for a single model.
`Plotter_Loop.ipynb`	Plots output of `Calc_Ind.ipynb` for all models specified.
`Plotter_Cumulative.ipynb`	Plots output of `Calc_Cumulative.ipynb` as combined plots for easier direct comparison
`Plotter_Sensitivity.ipynb`	Plots sensitivity analysis output of `Calc_Cumulative.ipynb` as a single plot with all configuration side by side.

Table 5 Plotting notebooks and their outputs.

All notebooks can be downloaded from the HighRes-Sea-Ice-Diagnostic-Toolset GitHub repository at https://github.com/CCarverUCT/HighRes-Sea-Ice-Diagnostic-Toolset (Last access: June 2025). The accompanying python environment file, `HRSIDT.yml`, is also provided in this repository.

3.1.2 Testing with CMIP6 data

Initial testing was completed with all historical CMIP6 models that had SIC datasets available on the Pangeo repository (Table 1; Last access: December 2024). All output metric plots of these 24 models include spatial plots of the cumulative, seasonal and monthly temporal mean, a time series plot of the monthly spatial mean values and corresponding seasonal and monthly distributions, and aggregated metric values are presented in Appendix B. Spatial plots of the cumulative temporal mean value for each configuration of the sensitivity analysis and heat mapped value plots of their respective aggregated values can be accessed in Appendix C. These are included for full representation of outputs generated throughout the development of this toolset. Specific analysis is not completed on any of the plots generated from processing of CMIP6 models in thesis because the focus is on high-resolution data.

3.1.3 Runtime and dataset size

The runtime of the `Calc_Cumulative.ipynb` script, defined in Sec 2.2, for all six HighResMIP models to be called from Pangeo, pre-processed, reprojected, and MEF metrics calculated was 188 seconds, about 31 seconds per model, when run over a 50 MB/s Wi-Fi connection. The metrics for 24 CMIP6 models were produced in 426 seconds, about 18 seconds per model, with the same network connection speeds. There are displayed in Table 5 below.

24 CMIP6 Models	6 HighResMIP Models
Completed 23 of 24 Completed 24 of 24 Done Total time take: 426.10 seconds	Completed 5 of 6 Completed 6 of 6 Done Total time take: 118.12 seconds

Table 6 Comparison of runtimes for the metric calculation for 24 CMIP and 6 HighResMIP models.

Export, of these calculated metrics, to the local device for all six HighResMIP models over the same Wi-Fi connection was completed in 376 seconds, about 63 seconds per model. The export of calculated metrics for all 24 CMIP6 models was completed in 252 seconds, about 11 seconds per model. Runtimes are impacted by the download speeds of the network being utilised by the users device.

24 CMIP6 Models
Done Total time taken: 252.17 seconds Frozen({‘models’: 24, ‘y’:332, ‘x’: 316, ‘season’:4, ‘month’: 12, ‘tdim’: 432})
6 HighResMIP Models
Done Total time taken: 375.94 seconds Frozen({‘models’: 6, ‘y’:332, ‘x’: 316, ‘season’:4, ‘month’: 12, ‘tdim’: 432})

Table 7 Comparison of the metric download runtimes for 24 CMIP and 6 HighResMIP models.

The size of the exported files for the 24 CMIP65 models and 6 HighResMIP models are 513.8MB and 128.5MB respectively, about 21.5MB per model.

3.2 HighResMIP MEF plots

All HighResMIP models available on Pangeo (Last access: December 2024) are run through the High-Resolution Sea Ice Diagnostics Toolset. The six models evaluated are: CMCC-CM2-HR4, CMCC-CM2-VHR4, GFDL-CM4C192, HadGEM3-GC31-HM, HadGEM3-GC31-LL, and HadGEM3-GC31-MM. These are all run with the NSIDC observational dataset, defined in Sec. 2.2.1, as the observational reference dataset and grid. The univariate metric calculated is MEF, defined in Sec. 2.2.5, using equations specified in Sec. 2.2.6 and 2.2.7. The GFDL-CM4C192 model has only one dataset available for analysis, deeming an intercomparison of different resolutions not possible. This model’s plots are included in Appendix A.1 for

reference and no further analysis is presented on them. The interpretation of the results is discussed in Chapter 4.

The results of the remaining five models are evaluated in two groups according to their respective base models, CMCC-CM2 and HadGEM3-GC31. The two CMCC-CM2 models vary in configuration with an increase in nominal atmosphere resolution from 100km (HR) to 25km (VHR). There is no change in nominal ocean resolution between the two configurations. The three HadGEM3-GC31 models vary in configuration with an increase in nominal atmosphere resolution from 250km (LM) to 100km (MM), and further to 50km (HM). An increase in nominal oceanic resolution is also present between 1° (LM) and 0.25° for both other models (MM and HM). These variations in component resolution are specified in Table-2.

All plot color mapping in this section follows the MEF categorization defined in Sec. 2.2.5 and are visualised according to the designated color defined below in Fig 3-3.

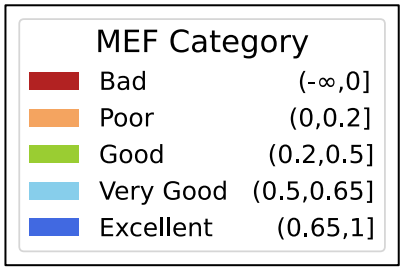


Figure 3-1 Colour mapping legend for the MEF metric categorisation. Applicable to all plots displayed in Section 3.

3.2.1 CMCC-CM2

The cumulative temporal mean MEF values, calculated with Eq. 1 (Sec. 2.2.6) and plotted in Fig. 3-2 below, show inconsistent localized improvement in model performance with an increase in atmosphere resolution (Table 2). Model performance is ‘bad’ in the Weddell Sea, adjacent to the Antarctic Peninsula for both models. The aggregated MEF value only differs by 0.01, no significant change is seen with the increase in nominal atmosphere resolution present between the two models.

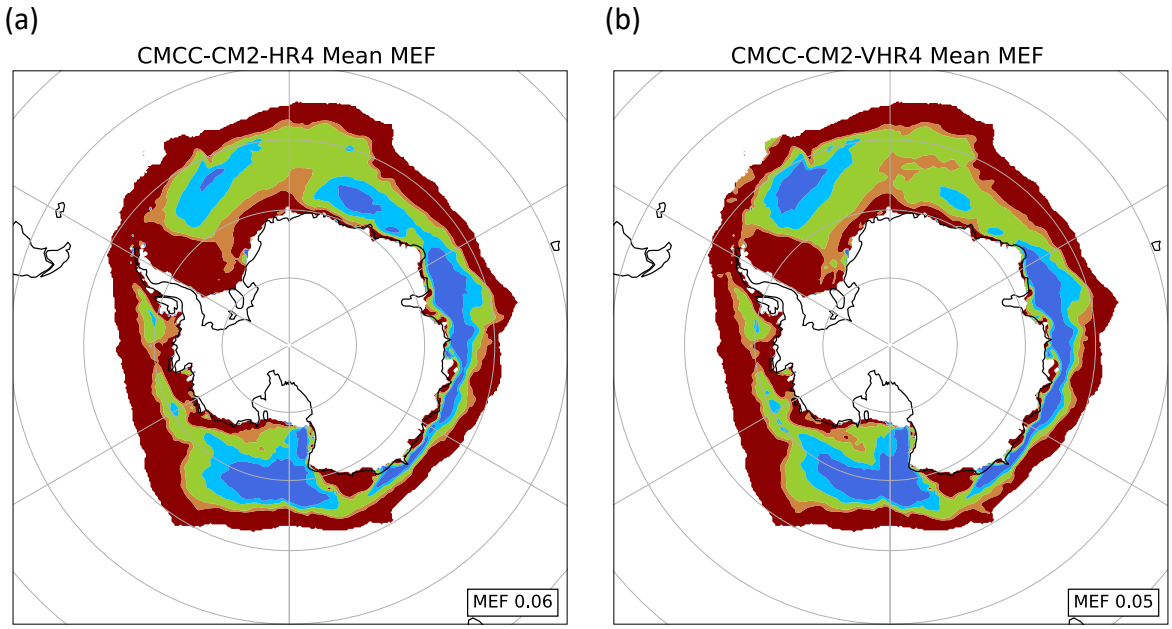


Figure 3-2 CMCC Cumulative temporal mean MEF

Models CMCC-CM2-HR4 (a) with 100km nominal atmosphere resolution and CMCC-CM2-VHR4 (b) with 25km nominal atmosphere resolution. Aggregated MEF values displayed in bottom right corner of each plot. Colour contours follow legend defined in Figure 3-1.

Grouping of seasonal spatial mean MEF values, calculated with Eq. 2 (Sec. 2.2.6) and plotted in Fig. 3-3, reveals reduced MEF values with the higher resolution model for season DJF while the other three seasons are near identical between both models. Performance is generally worse nearer the continent for DJF and MAM while 'excellent' performance is seen surrounding most of the continent in JJA and SON. 'bad' performance is seen throughout all seasons in the Weddell Sea.

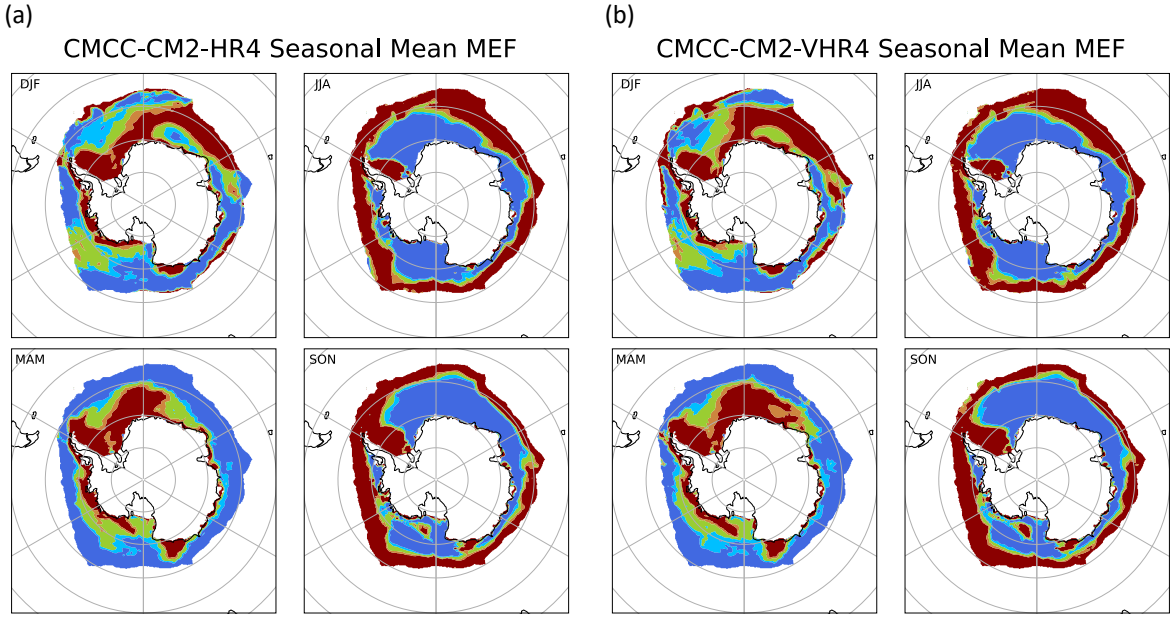


Figure 3-3 CMCC Seasonal temporal mean MEF

Seasonal temporal mean MEF values of models CMCC-CM2-HR4 (a) with 100km nominal atmosphere resolution and CMCC-CM2-VHR4 (b) with 25km nominal atmosphere resolution. seasonal temporal mean MEF values. Season specified in each plot's top left corner. Colour contours follow legend defined in Figure 3-1.

Monthly climatological MEF values for the CMCC-CM2 model, calculated with Eq. 3 (Sec. 2.2.6) and plotted in Fig. 3-4, reveal all months other than January and December to have similar performance across the grid between both atmosphere resolutions, there is some improvement in performance in the Weddell Sea and the Ross Sea regions, but no overall pattern is apparent. For the model with increased atmosphere resolution, January demonstrates a noticeable reduction in performance in the Weddell Sea and Western Indian Ocean regions, where slight improvement is seen in the Ross Sea. December appears to be an outlier with 'bad' MEF performance across most of the domain. There is a gradual shift in areas of 'excellent' versus 'bad' performance as the months progress starting with February exhibiting 'excellent' performance away from the continent and 'bad' performance near continent. This near continent 'bad' performance expands for March and April before turning into an internal ring of 'bad' performance for May and June. July through October have 'excellent' performance for a considerable region near continent with 'bad' performance at the external ring of sea ice. November and December see rapid decay in performance. The shifting regions of 'bad' performance seem to align with regions where the MIZ would be expected, SIC values between 15% and 80%. This will be further investigated by the sensitivity analysis results in Sec. 3.3.1.

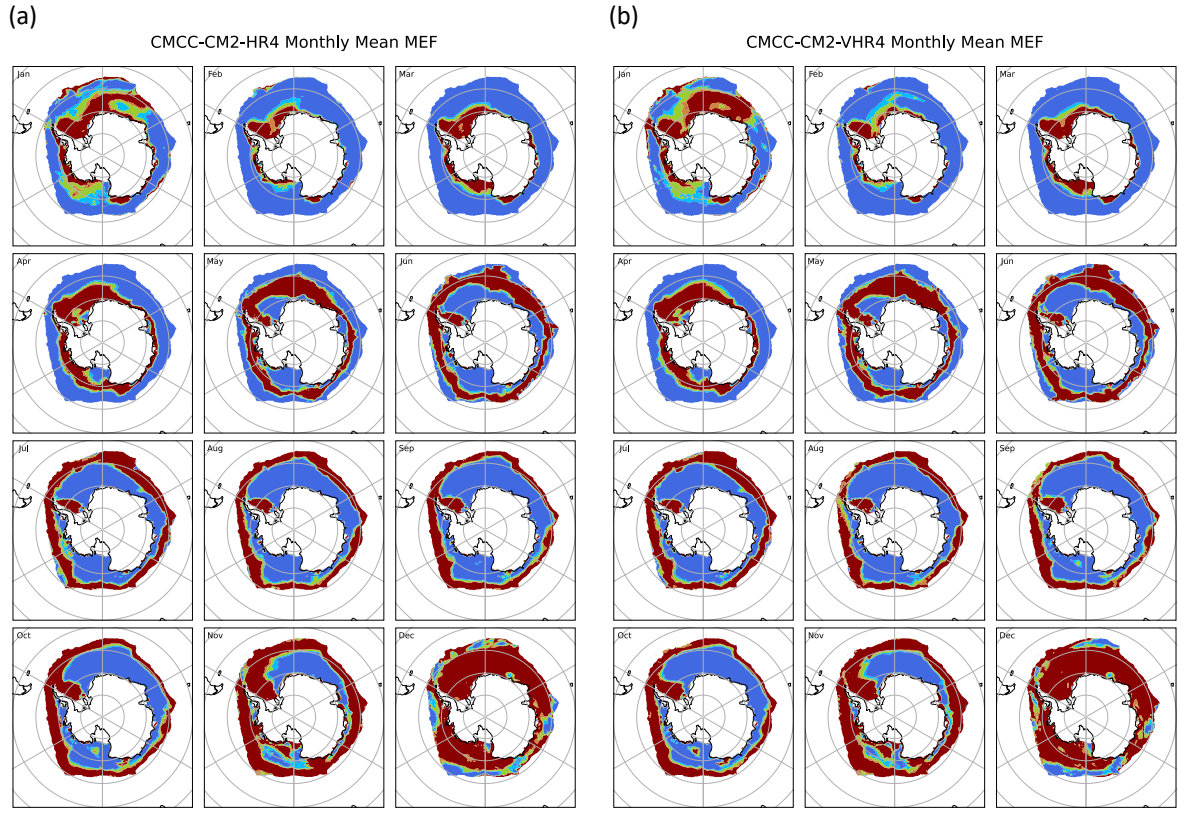


Figure 3-4 CMCC Monthly climatological MEF

Monthly temporal mean MEF values of models CMCC-CM2-HR4 (a) with 100km nominal atmosphere resolution and CMCC-CM2-VHR4 (b) with 25km nominal atmosphere resolution. Month specified in each plot's top left corner. Colour contours follow legend defined in Figure 3-1.

The monthly spatial mean MEF values, calculated with Eq. 4 (Sec. 2.2.7) and plotted in Fig 3-5, show no evident trend in model performance for either resolution. There is considerable seasonal variation in MEF values with modulation between ‘good’ to ‘excellent’ and ‘bad’ values seen.

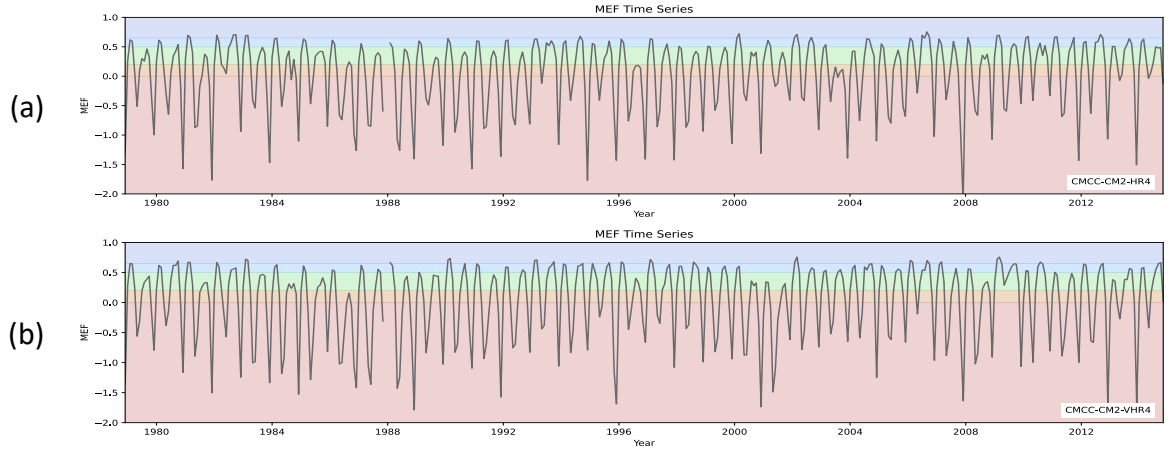


Figure 3-5 CMCC Monthly spatial mean MEF time series

Time series of monthly spatial mean MEF for models CMCC-CM2-HR4 (a) with 100km nominal atmosphere resolution and CMCC-CM2-VHR4 (b) with 25km nominal atmosphere resolution. Colour overlay follows legend defined in Figure 3-1.

Visualization of the seasonal and monthly distribution of spatial monthly mean MEF values, calculated with Eq. 4 (Sec. 2.2.7) and plotted in Fig. 3-6, show variation in the impact of increased horizontal resolution. The model with increased atmosphere resolution, Fig. 3-8b, shows both an increase in the count of ‘bad’ and ‘very good’ MEF performance for the season of DJF. JJA sees a loss of any instances of performance better than ‘good’. SON centers more on ‘good’ performance losing ‘very good’ and ‘bad’ instances. DJF spreads to extremes with the absence of ‘good’ and ‘very good’ instances and the development of ‘excellent’ instances appearing with the high-resolution run. The monthly distribution shown in Fig. 3-8b,d shows general performance decay across all months with increased resolution. Seasonal variability is more apparent here with May, June, November and December all having more than half of their occurrences with ‘bad’ performance. February, March, September and October have the best performance and May, June, November and December have the worst performance across both models.

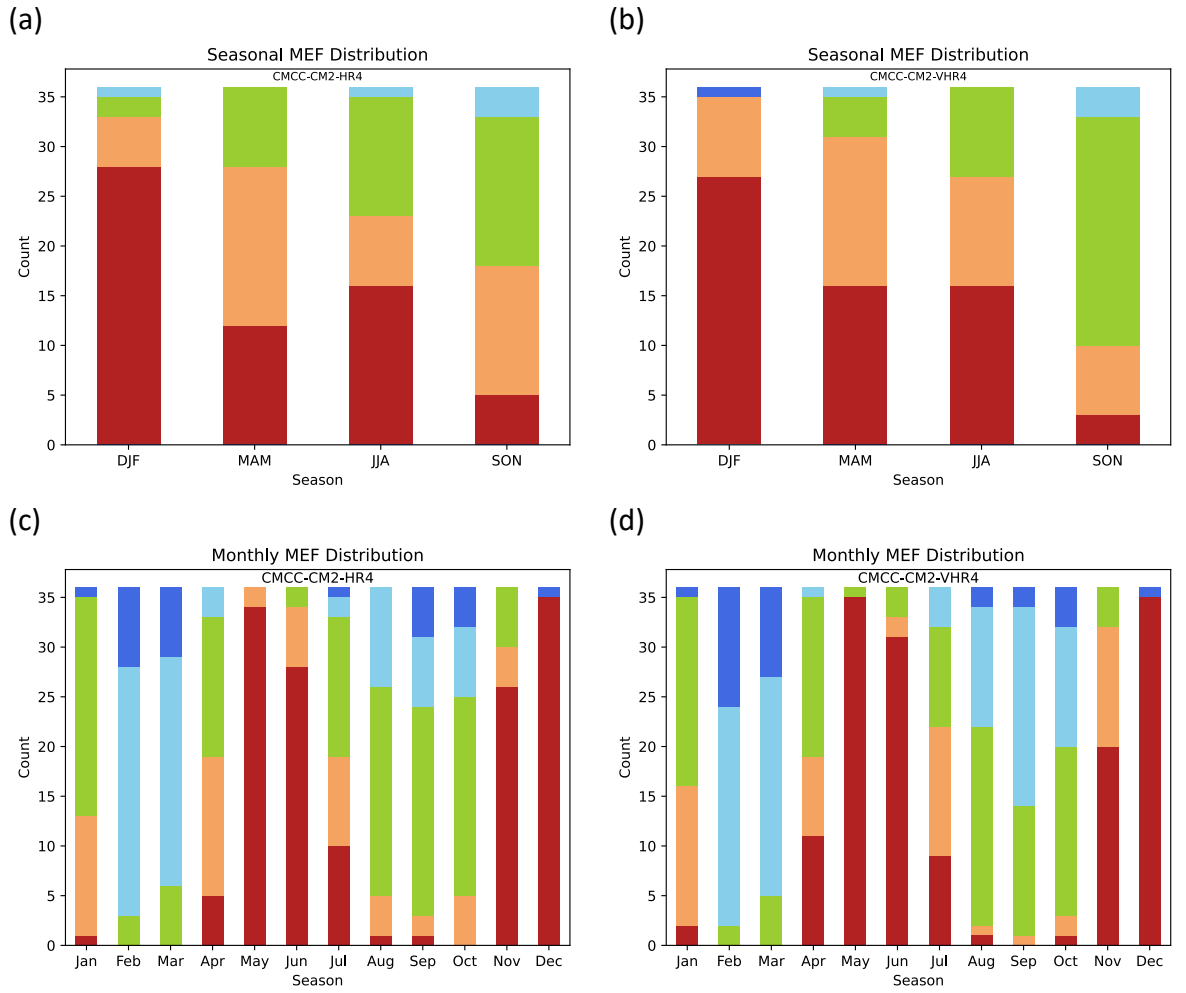


Figure 3-6 CMCC Spatial mean MEF distribution

Seasonal and monthly mean spatial MEF values distribution for models CMCC-CM2-HR4 (a,c) with 100km nominal atmosphere resolution and CMCC-CM2-VHR4 (b,d) with 25km nominal atmosphere resolution. Colour categorisation follows legend defined in Figure 3-1.

3.2.2 HadGEM3-GC31

The cumulative spatial mean MEF values, calculated with Eq. 1 (Sec. 2.2.6) and plotted in Fig. 3-7, for the three different resolution configurations, LL (a), MM (b), and HM (c), of the HadGEM3-GC31 model show marginally increased performance with the middle resolution and slight performance decay with the highest resolution version. The model performance in the southwestern Weddell Sea along the Antarctic Peninsula is 'bad' for all resolutions, as it is for the East Antarctic and Bellingshausen regions bordering the Ross Sea. The aggregated MEF values indicated a negligible difference in performance across all three with only a +0.02 improvement from LL to MM and a smaller +0.01 improvement LL to MM. Visually and quantitatively the middle resolution model marginally has the better performance. Modelling of the marginal ice zone, where the lowest SIC is expected, has uniform 'bad' performance.

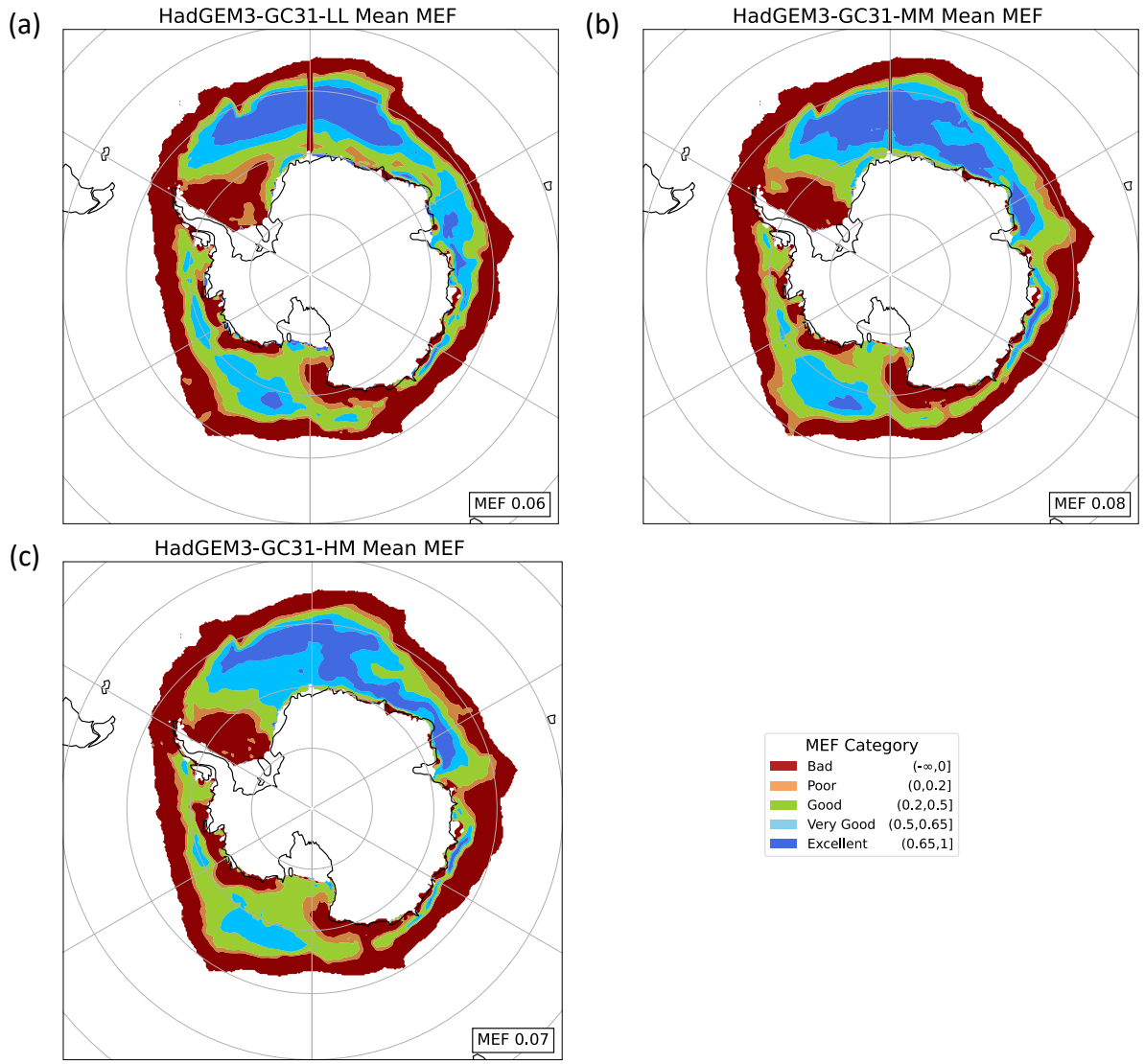


Figure 3-7 HadGEM3 Cumulative temporal mean MEF
(a) HadGEM3-GC31-LL, (b) HadGEM3-GC31-MM, and (c) HadGEM3-GC31-HM. Aggregated MEF values displayed in bottom right corner of each plot.

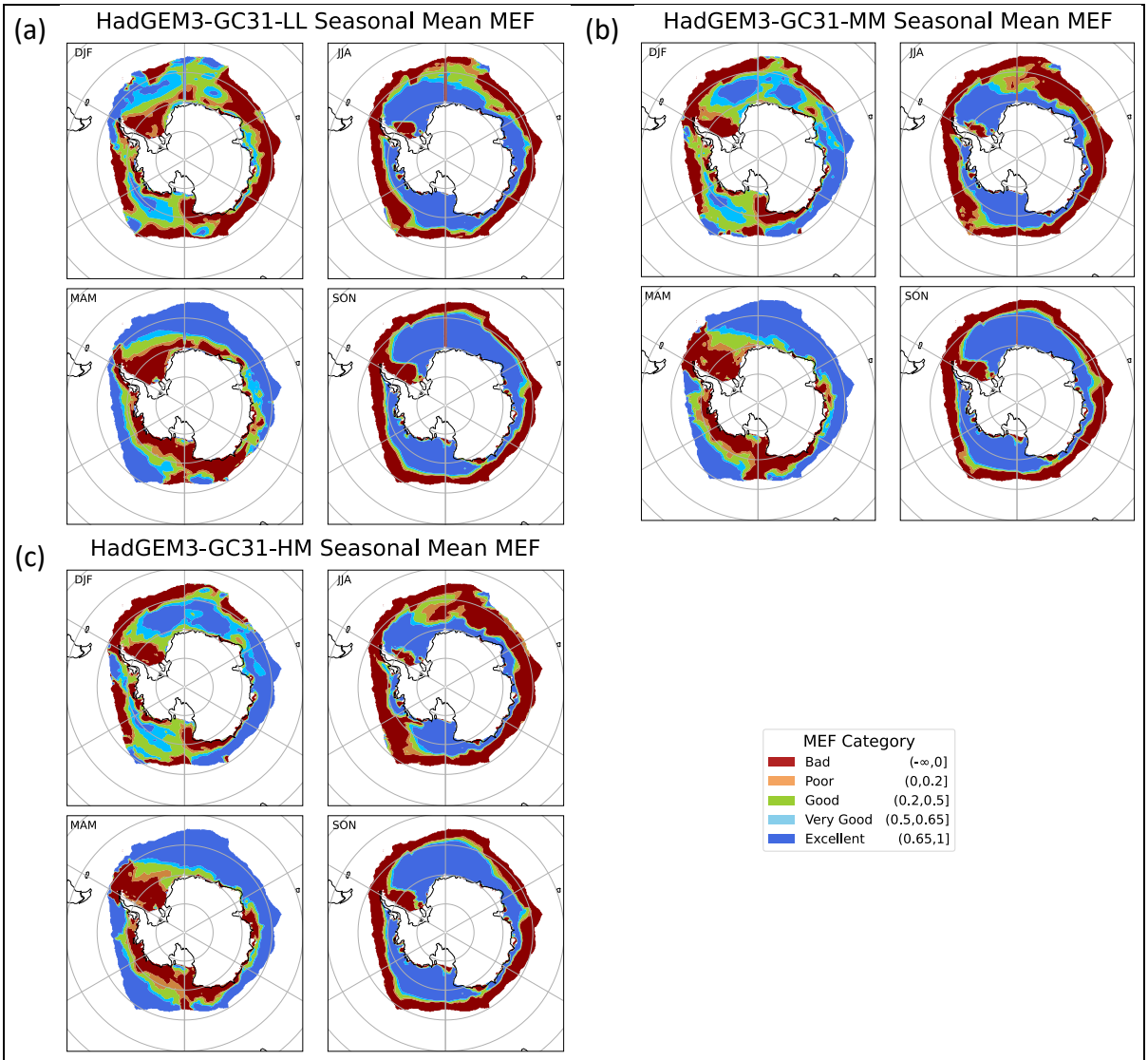


Figure 3-8 HadGEM3 Seasonal temporal mean MEF

For models HadGEM3-GC31-LL (a) with 1º ocean and 250km atmosphere resolution, HadGEM3-GC31-MM (b) with 0.25º ocean and 100km atmosphere resolution and HadGEM3-GC31-HM (c) with 0.25º ocean and 50km atmosphere resolution. MEF values with season specified in each plot's top left corner.

The grouping of seasonal spatial mean MEF values, calculated with Eq. 2 (Sec. 2.2.6) and plotted in Fig. 3-8, reveals distinct trends for each season. DJF exhibits slightly improved performance with increased model resolution identified by larger areas of 'excellent' MEF values between longitudes 60°E and 180°E. This increased performance is particularly noticeable around the edge of the sea ice area between the lowest resolution and the other two resolutions which have little difference between them. This aligns with an increase in nominal ocean resolution specified in Table 2. MAM shows varied change in performance with increased ocean resolution and negligible change with increased atmosphere resolution from MM to HM. JJA demonstrates deterioration in performance with increased resolution. SON shows no significant change in overall performance across all resolutions. 'Bad' model performance is seen throughout all seasons in the Weddell Sea.

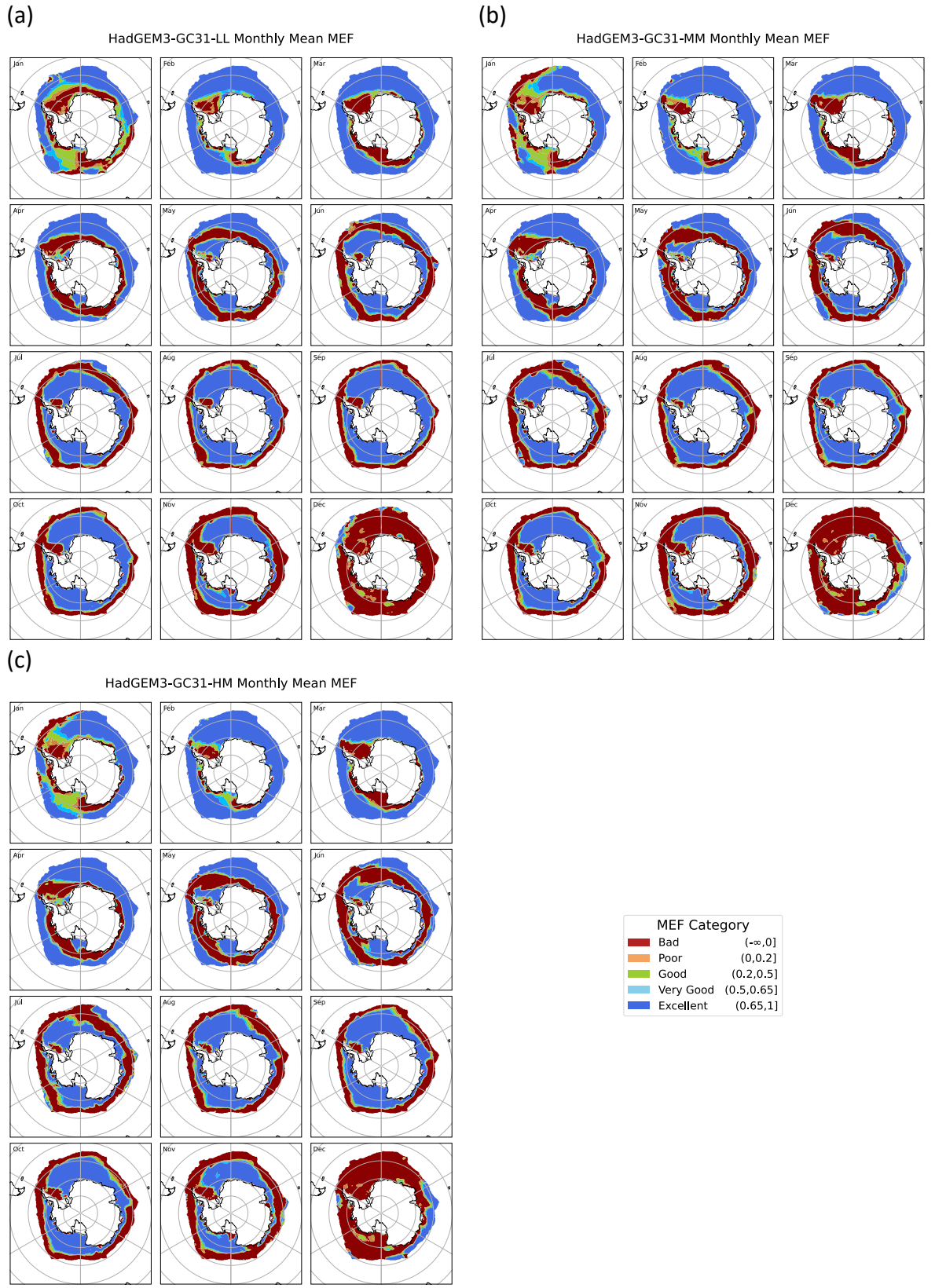


Figure 3-9 HadGEM3 Monthly climatological mean MEF

For models HadGEM3-GC31-LL (a) with 1º ocean and 250km atmosphere resolution, HadGEM3-GC31-MM (b) with 0.25º ocean and 100km atmosphere resolution and HadGEM3-GC31-HM (c) with 0.25º ocean and 50km atmosphere resolution. Month specified in each plot's top left corner.

Monthly spatial mean MEF values for the HadGEM3-GC31 model, calculated with Eq. 3 (Sec. 2.2.6) and plotted in Fig. 3-9, show all months, other than January and Dec, to have similar performance across the grid between resolutions, there is some decay in performance in localised regions, but no overall trend is apparent. January demonstrates noticeable reduction in performance of the model with increased resolution in a larger area where some improvement is seen in others. December appears to be an outlier with 'bad' MEF performance across most of the grid with performance improvement with increased resolution. There is a gradual shift in areas of 'excellent' versus 'bad' performance as the months progress starting with February exhibiting 'excellent' performance away from the continent and 'bad' performance near continent. This near continent 'bad' performance expands for March and April before turning into a detached ring of 'bad' performance for May, June and July. August through November have 'excellent' performance for a considerable region near continent with 'bad' performance at the external ring of sea ice. December displays a drastic decay in performance. January sees a sudden shift back to near continent modelling having 'bad' performance and an external ring of 'excellent'. Focusing on the Weddell Sea it can be seen that the model exhibits 'bad' performance throughout all months and resolutions.

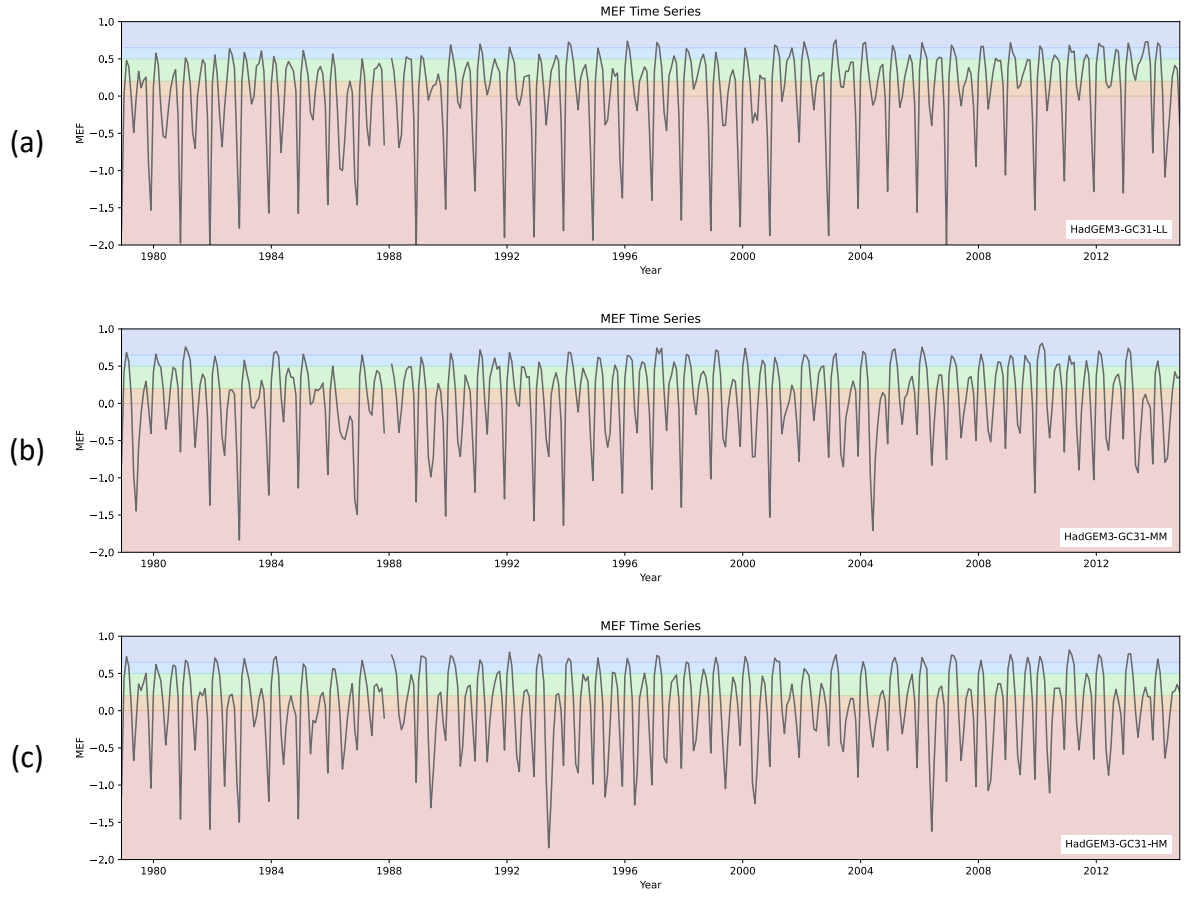


Figure 3-10 HadGEM3 Monthly spatial mean MEF time series
For models HadGEM3-GC31-LL (a) with 1º ocean and 250km atmosphere resolution, HadGEM3-GC31-MM (b) with 0.25º ocean and 100km atmosphere resolution and HadGEM3-GC31-HM (c) with 0.25º ocean and 50km atmosphere resolution.

The monthly spatial mean MEF values, calculated with Eq. 4 (Sec. 2.2.7) and depicted in Fig. 3-10(a-c), show no evident trend throughout the timespan for model performance across any of the model resolutions. There is considerable seasonal variation in MEF values with modulation between 'good' to 'excellent' and 'bad' values seen. Marginal change in seasonal performance behavior is seen with the increased magnitude of negative MEF values throughout the entire time series with increased resolution.

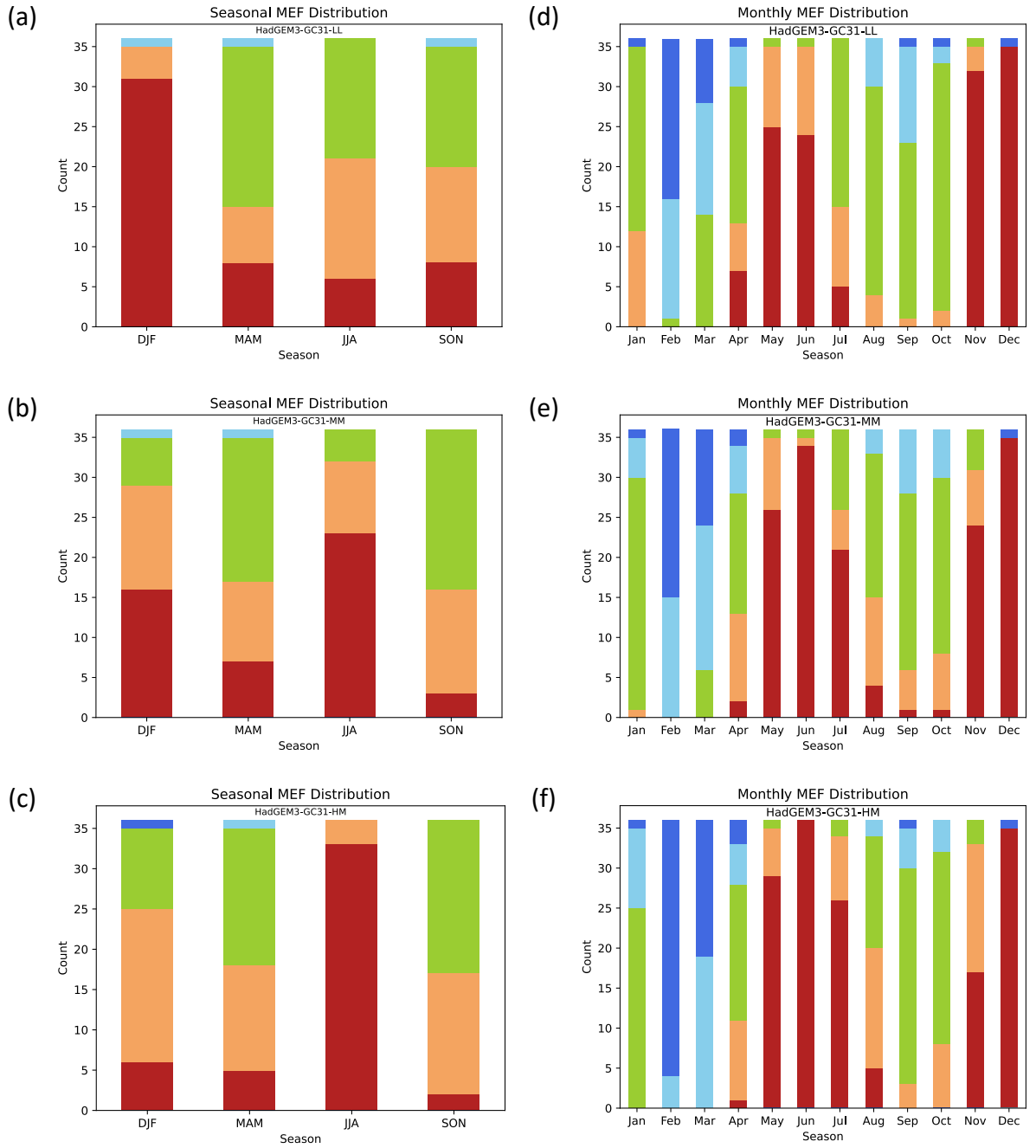


Figure 3-11 HadGEM3 Spatial mean MEF distribution

Seasonal and monthly mean spatial MEF distribution for models For models HadGEM3-GC31-LL (a,d) with 1º ocean and 250km atmosphere resolution, HadGEM3-GC31-MM (b,e) with 0.25º ocean and 100km atmosphere resolution and HadGEM3-GC31-HM (c,f) with 0.25º ocean and 50km atmosphere resolution. Colour categorisation follows legend defined in Figure 3-1.

Visualization of the seasonal and monthly distribution of spatial monthly mean MEF values, for the HadGEM3-GC31 model, calculated with Eq. 4 (Sec. 2.2.7) and plotted in Fig. 3-11, reveal variation in the impact of increased horizontal resolution. Seasonal distributions, shown in Fig. 3-11 (a-c), reveal that JJA season distribution experiences an increased occurrence of 'bad' MEF values. Both MAM and SON see a slight improvement of MEF performance distribution while DJF has significant improvement with the increase of resolution. Monthly distributions, shown in Fig. 3-11 (d-f), reveals general improvement with increased resolution for the months of January-March and November. April – August experiences moderate deterioration in performance with increased resolution. September and October exhibit the best performance distribution with the lower resolution followed by the highest resolution model and the middle resolution performing the worst. December is consistent across all model resolutions with all but one occurrence being 'bad'.

3.2.3 Comparison of aggregated MEF values

The aggregated MEF values, calculated with Eqs. 5-7 (Sec. 2.3.4) and displayed in Fig. 3-12 summarize the features seen in the above results for all the models. This figure also includes the results of the GFDL-CM4C192 model that were not analysed in detail.

Figure 3-12 Aggregated MEF values

Annual, seasonal, and monthly aggregated MEF values for all six HighResMIP models displayed, and heat mapped according to categorical performance colours defined in Figure 3-1.

The CMCC-CM2 model has varied results for MEF performance improvement between the two different resolution runs, HR4 and VHR4. Annually, there is a negligible decrease in performance (-0.01) for the increased resolution model. DJF sees no change, MAM and JJA see small deterioration (-0.07, -0.03) and SON sees small (+0.06) improvement. Monthly, January and April – June show a reduction in performance (range of -0.03 to -0.19) February, March, and August – December saw small improvements (range of +0.01 to +0.07) in performance. July saw no change.

The HadGEM3-GC31 model has similar minimally varied results across model resolutions with the middle resolution model, MM, having a +0.02 improvement over the lower

resolution, LL, and a +0.01 improvement over the higher resolution model (HM). This pattern is replicated in the seasonal values only for MAM. MEF values for both DJF and SON increase with model resolution. JJA decreases with increased model resolution. Monthly MEF values show improvement with increased resolution for January – April, November and December. MEF values for May – August decrease with increased resolution. September and October decrease in performance from low to medium resolution but have a negligible change with the increase to high resolution.

The GFDL-CM4C192 model shows seasonal and monthly fluctuations, matching all other models with January – April and August - October having 'good' to 'excellent' performance while May, Jun, November, and December display 'bad' performance. Overall, this model demonstrates the best performance of all HighResMIP models with a 'good' aggregate MEF value of 0.25.

3.3 Sensitivity analysis

All sensitivity analysis configurations, defined in Sec. 2.2.9, are run for each HighResMIP model. The cumulative temporal mean MEF and aggregated MEF values of these configurations, in addition to the default configuration, are plotted below in Figures 3-13 and 3-14. Due to a high level of similarity observed in the results across different resolutions of the same model, only outputs from a single resolution for each model CMCC-CM2 and HadGEM3-GC31 are included here. The results generated for other resolutions of these two models as well as the GFDL-CM4C192 model are included in Appendix A.2 for reference. The complementing monthly spatial mean MEF value distribution stacked bar charts, for both seasonal and monthly groupings, are included in Appendix A.3 for completion.

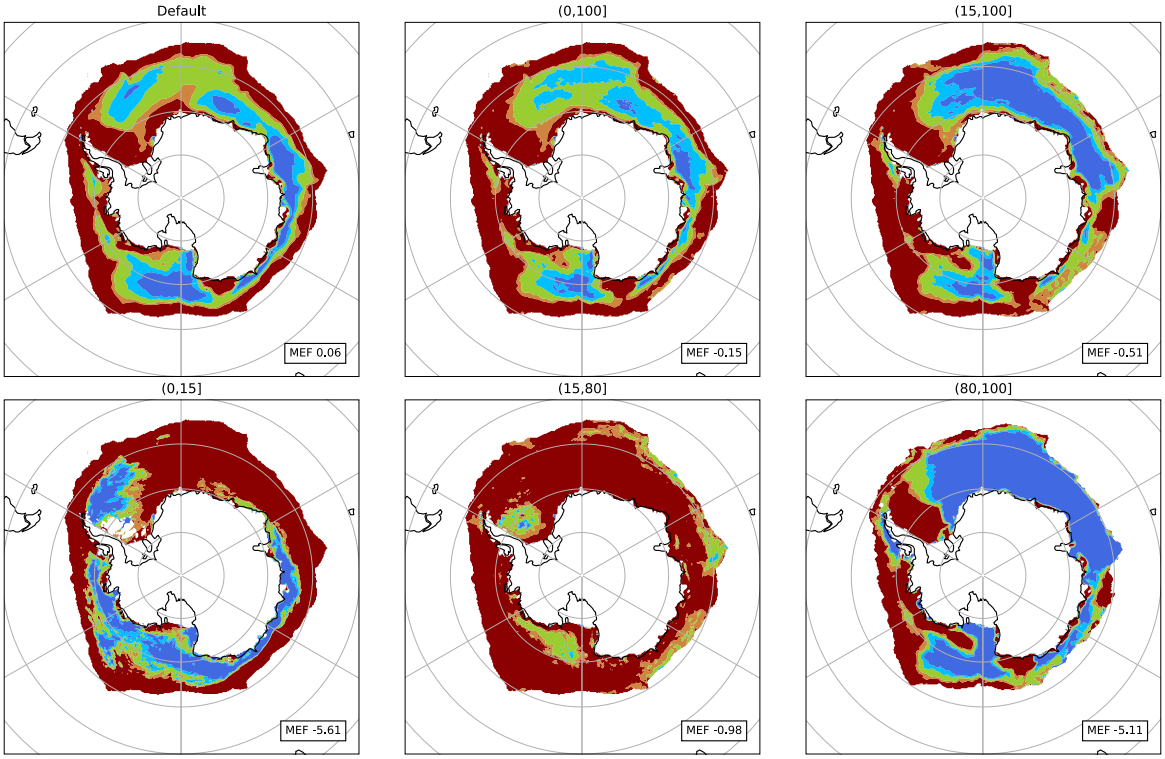
3.3.1 CMCC-CM2

The CMCC-CM2 model exhibits reduced performance with the exclusion of 0% sea ice concentration, (-0.21 change) and more so with the exclusion of all values less than or equal to 15% (-0.57 change). Though a 'bad' aggregated MEF value is observed for both configurations, the spatial plots of the temporal mean values (Fig. 3-13(a)) reveal increased areas of 'excellent' values with (15,100] concentration range having significantly more 'excellent' area relative to the default. This implies large magnitude in the 'bad' metric values. For (0,15] concentration range the model shows some areas of 'excellent' performance but overall, the 'bad' areas are extensively more negative resulting in an aggregated value of -5.61. The (15,80] concentration range shows widespread 'bad' performance across the entire plot. The aggregated MEF value of -0.98 indicates few extreme values, rather a more uniform 'bad' performance. The (80,100] concentration range shows overall 'excellent' performance across much of the grid area though the 'bad' region departs significantly from the observational mean as the cumulative compressed MEF value is -5.19.

Considering the aggregated mean MEF values of Fig. 3-13(b) the groupings of (0,100] and (15,100] show improved model performance during JJA and SON, and significantly reduced values for DJF and MAM. Monthly improvement is seen only in June – November, a time

when concentration values are expected to be higher. Regarding (15,80] concentration range the model performs significantly worse than the mean observational values for all months other than September and October where it nearly matches the mean with an aggregated MEF value of -0.03. Extremely 'bad' (< -10) performance is present for January – March, when the lowest SIC is expected.

(a)



(b)

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.06	-0.11	0.08	0.06	0.22	0.23	0.60	0.58	0.19	-0.52	-0.38	0.15	0.39	0.45	0.39	-0.20	-1.19
(0,100]	-0.15	-0.60	-0.56	0.23	0.31	-0.20	-0.11	-0.15	-0.66	-0.86	-0.16	0.34	0.51	0.54	0.47	-0.07	-1.50
(15,100]	-0.51	-1.61	-1.20	0.36	0.39	-1.29	-1.46	-1.17	-1.52	-0.91	0.05	0.46	0.58	0.59	0.52	0.05	-2.08
(0,15]	-5.61	0.13	-0.02	-10.20	-12.25	0.45	0.61	0.53	0.19	-0.77	-2.62	-8.24	-19.75	-21.52	-10.34	-4.88	-0.70
(15,80]	-0.98	-0.74	-1.72	-1.10	-0.37	-0.18	-0.10	-0.36	-1.43	-3.38	-2.33	-0.78	-0.18	-0.03	-0.03	-1.05	-1.97
(80,100]	-5.11	-14.34	-7.54	0.67	0.60	-13.50	-26.74	-20.64	-1.96	-0.02	0.55	0.71	0.76	0.75	0.67	0.37	-2.43

Figure 3-13 CMCC-CM2-HR4 sensitivity analysis plots

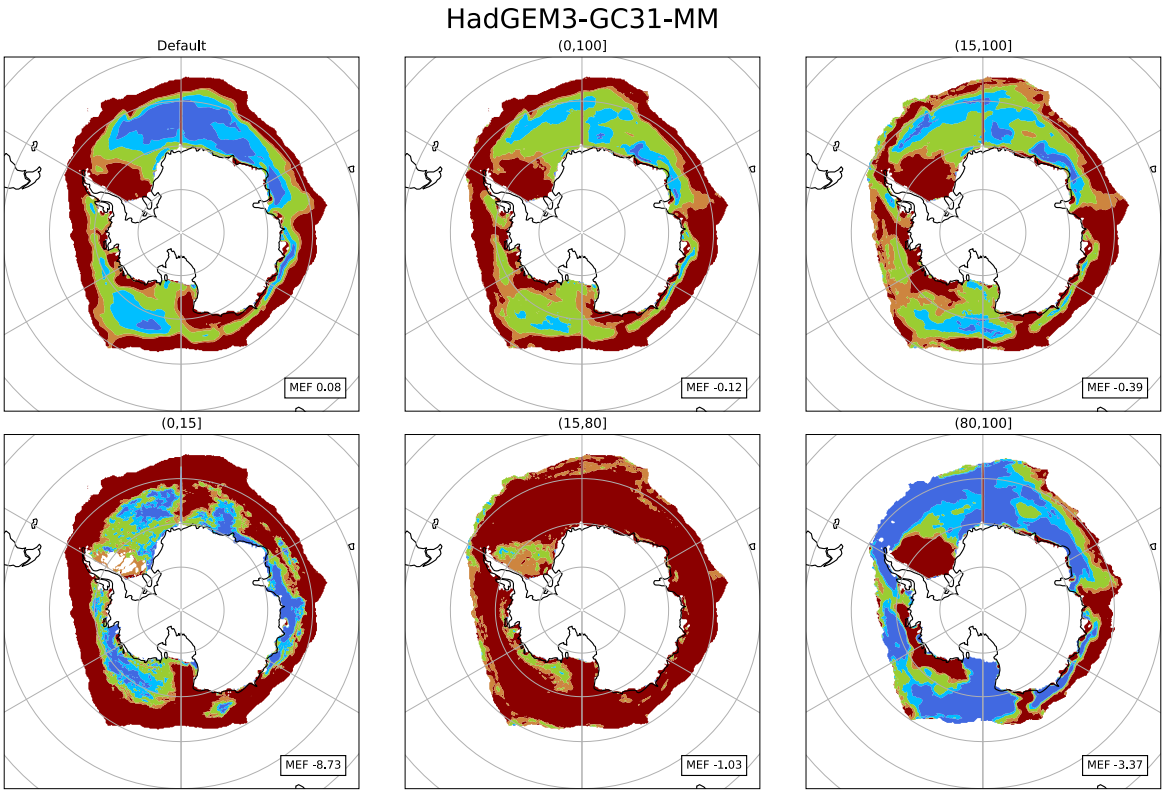
Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations of different sea ice concentration percentage groupings specified in their respective plot titles. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat map colouring are mapped to categorical performance colours defined in Figure 3-1.

3.3.2 HadGEM3-GC31

Across all sensitivity analysis configurations the HadGEM3-GC31 model (Fig. 3-14(a)) exhibits similar metric performance behaviour to that seen in the CMCC-CM2 model. Reduced aggregated MEF values are observed for the (0,100] concentration range (-0.20 change) and for the (15,100] concentration range (-0.47 change). This does deviate from behaviour seen with the CMCC-CM2 model with decreased area of 'excellent' performance seen across both these concentration ranges relative to the default range of concentrations, though overall area of 'good; or better increases. Areas of 'excellent' performance are present for concentration range of (0,15] but the aggregated MEF has a large negative value of -8.73, again indicating the areas of 'bad' performance have negative values of large magnitude. The same (15,80] concentration range shows uniform 'bad' performance across the entire plot though the aggregated MEF value is only -1.03. The (80,100] concentration range shows 'excellent' performance across much of the grid area though the few 'bad' regions depart significantly from the observational mean as the cumulative compressed MEF value is -3.37. The model prediction for the Weddell Sea shows 'bad' performance for this concentration range.

The aggregated mean MEF values, Fig. 3-14(b) behaviourally match those of the CMCC-CM2 model with slight variation in the length of time periods with better than 'bad' performance observed for each of the (0,15], (15,80], and (80,100] concentration ranges. The model has 'good' or better performance for (80,100] concentrations for July-November when more sea ice is expected and significantly struggles in this concentration range for January - April when the least sea ice is expected. The performance is flipped for concentrations (0,15] with 'good' or better performance seen January - April and significantly 'bad' performance seen July-November. For concentrations (15,80] the model is uniformly 'bad' March–December, only performing marginally better than the observational mean in January and Feb, 0.03 and 0.07 respectively.

(a)



(b)

HadGEM3-GC31-MM

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.08	0.02	0.21	-0.10	0.19	0.39	0.66	0.58	0.32	-0.28	-0.51	-0.05	0.25	0.36	0.32	-0.12	-1.02
(0,100]	-0.12	-0.26	-0.64	0.00	0.41	0.04	0.07	-0.21	-0.74	-0.97	-0.45	0.09	0.37	0.48	0.50	0.26	-0.91
(15,100]	-0.39	-0.77	-1.44	0.10	0.55	-0.47	-0.92	-1.16	-1.84	-1.33	-0.37	0.20	0.46	0.57	0.62	0.47	-0.91
(0,15]	-8.73	0.01	0.27	-10.99	-24.03	0.35	0.60	0.44	0.37	-0.00	-1.36	-7.73	-23.86	-38.17	-25.62	-8.30	-0.93
(15,80]	-1.03	-0.45	-1.85	-1.39	-0.40	0.03	0.07	-0.50	-1.53	-3.53	-2.69	-1.06	-0.44	-0.22	-0.18	-0.80	-1.45
(80,100]	-3.37	-8.44	-6.36	0.41	0.82	-6.02	-19.09	-16.07	-2.48	-0.53	0.12	0.45	0.67	0.78	0.84	0.85	0.09

Figure 3-14 HadGEM3-GC31-MM sensitivity analysis plots

Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations of different sea ice concentration percentage groupings specified in their respective plot titles. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat map colouring are mapped to categorical performance colours defined in Figure 3-1.

Chapter 4
Discussion and Conclusions

4.1 New metric toolset

With the described limitations and future compatibility issues of SITool (Sec. 2.1), it was decided to develop a new toolset to provide further analysis capabilities on the impact of horizontal resolution on climate model performance with regards to Antarctic sea ice. The newly developed toolset resolves many challenges faced when working with high-resolution data including large dataset sizes and extensive computational power requirements. Visual outputs from the toolsets provide useful insights into the impact of horizontal resolution with the cumulative temporal mean plots (Figs. 3-2 and 3-7) immediately highlighting areas of different performance levels.

While the cumulative temporal mean visualisations provide a general overview of model performance, the grouping and visualisation of the seasonal and monthly temporal means (Figs. 3-3, 3-4, 3-8, and 3-9) reveal specific strengths and weaknesses of the respective model. The importance of visualising these groupings is seen in the two models assessed, with revealed performance weakness in the MIZ, which is only present across the monthly visualisations. This weakness is not apparent in the cumulative mean as it is averaged out over all months. The weakness of both models in reproducing December sea ice also becomes apparent, emphasizing the tendency for models to struggle across lower SIC values. The time series visualisations of spatial mean metric values (Figs. 3-5 and 3-10) provide a straightforward way to identify model performance decay over time. Though no trends are present in the two HighResMIP models, decay in performance was observed in the initial testing with CMIP6 models. Distributions of the seasonal and monthly groupings of these values (Figs. 3-6 and 3-11) give insight into the performance trends in these respective groups. Aggregated metric values (Fig. 3-12) provide an initial notion of overall performance difference between multiple models.

All outputs of the toolset configured to produce the MEF metric for the defined HighResMIP plots indicate negligible difference in model performance with increased ocean and/or atmosphere horizontal resolution. This aligns with conclusions made by Selivanova *et al*. (2024). Furthermore, with Antarctic sea ice extent at a minimum in February and a peak in September, these two months are often used as representative months in sea ice modelling as practiced by Selivanova *et al*. (2024) and Lin *et al*. (2021). However, the results of the MEF metric across all models indicate that these months are the best months represented by the evaluated models, whereas the transitional seasons of May/June and November/December are very poorly represented by all models. This aligns with the notion of models recreating the inflection points of sea ice extent growth/recession and struggling with the transitional months. Taking this model behaviour into consideration, it is important to acknowledge the

performance bias that may be present from the selection any two months as indicators of overall model performance.

It is necessary to note that aggregated metric values produced by the toolset are only a general representation of model performance and do not provide significant insight into the specific characteristics of the model. This is due to the spatial and temporal aggregation of values which smooths out many variations in values greatly reducing the ability to detect any spatial or temporal trends. It must also be noted that these can be influenced by negative metric values of large magnitude. With the MEF metric the resulting range is a maximum of 1, while there is no defined minimum value, leading an aggregated value to be more negatively skewed. It is suggested that aggregated values should not be used as a definitive indicator of model performance, but as general guidelines when, for instance, comparing versions in the model development phase. Another limitation of this initial development is the exclusion of occurrence weighting when calculating spatial and aggregate mean metric values. A metric value at a discrete grid point that has a valid observational SIC value for all months contributes equally to the mean metric value as a metric value of a discrete grid point that only has a valid observational SIC value for some months. This can result in grid points with very poor performance across a lower occurrence of observational values having a disproportionate influence on the spatial and aggregate means compared to a point with complete observational occurrence.

4.2 Sensitivity analysis

Though sensitivity analyses showed negligible change between different model resolutions, of the two models reviewed, other insights were derived from the results (Sec. 3.3). Sensitivity analyses have revealed consistent shortcomings across all evaluated models in simulated SIC in the marginal ice zone (15-80%). In contrast SIC values between 80-100%, which are typically considered pack ice, sees ‘excellent’ model performance across most of the region for all models analysed. This shows that this feature of the toolset can be utilised to assess the development of any given model regarding specific ranges of the variable of interest. The five configurations demonstrated here can be modified as necessary to inspect desired range of values. The better a model’s performance is defined regarding sea ice of different groupings, the more it can be understood and improved upon in future iterations.

4.3 Pangeo and cloud-ready data

The toolset provides significant improvement in workflow efficiency of high-resolution model dataset analysis, evident with the run time observed (Sec. 3.1.3). The runtime per model should scale linearly as each model is projected onto the same reference grid. Should a higher resolution observational reference grid be used the processing time can be expected to increase while still providing substantial time saving. This is a result of the successful integration of a remote repository, Pangeo, and use of cloud ready data and cloud computing capacity. The toolset is also structured for easy modification to enable intake from any user defined remote repository provided the correct address. Runtimes for calculation and local

writing of all relevant metrics for multiple high-resolution models are completed in less time that a single high-resolution dataset could typically be downloaded. The elimination of requirement for local download, expansive storage and significant computing power, traditionally required of high-resolution datasets, provides greatly increased accessibility for all climate model datasets, both high and low resolution. With this functionality, users no longer need to be restricted by local storage capacity or computational power nor be slowed by extensive run times.

4.4 Conclusions and future development

The toolset developed and presented in this thesis is intended to enable the evaluation of a model's capability to predict Antarctic sea ice through the calculation and visualisation of any univariate model performance metric, provided a sufficient observational reference dataset. The aim was to produce a flexible operational tool that overcomes the issues of the tools available in the literature, and that can be used in the future to support scientific understanding. The MEF metric implemented here, for testing and demonstration, is one of many univariate metrics compatible with the toolset. The toolset also demonstrates how cloud ready data, remote repositories and cloud computing can be utilised effectively to enable the analysis of high-resolution climate models with significantly reduced local computational capacity requirements and overall runtimes. The initial output from running climatological SIC data of six HighResMIP models through the toolset is presented and the preliminary analysis conducted establishes the efficacy of the toolset to reveal insights regarding the impact of increased horizontal ocean and/or atmosphere resolution. Further analysis of the results is beyond the scope of this minor-dissertation and left to the intended future users of the developed toolset to further explore the inner workings of sea-ice components of climate models.

The toolset is intended to provide a backbone from which much more analyses can be produced and there is considerable potential for this toolset to be expanded upon. Potential expansions include the incorporation of analysis of other key sea ice characteristics and application to Arctic sea ice and associated variations in behaviour. Furthermore, occurrence weighted spatial averaging can also be explored to produce more meaningful aggregated metric values. Regardless of further development the toolset, in its current state, provides extensive flexibility across different univariate metrics, observational reference datasets, remote repositories, and sensitivity analysis configurations. Considering this adaptability and the accompanying thorough visualisation structure it is intended that the toolset be utilised to further the analysis and understanding of the role that increased horizontal resolution has on the performance of climate models when predicting Antarctic sea ice. This is with the ultimate aim that his aids in the development of climate models as a whole, improving the overall understanding of the world around us.

References

Bentsen, M., Oliviè, D.J.L., Seland, Ø., Toniazzo, T., Gjermundsen, A., Graff, L.S., Debernard, J.B., Gupta, A.K., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.8040.

Bethke, I., Wang, Y., Counillon, F., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H.R., Chiu, P.-G., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.10894.

Budyko, M.I. 1969. The effect of solar radiation variations on the climate of the Earth. Tellus A: Dynamic Meteorology and Oceanography. (January). DOI: 10.3402/tellusa.v21i5.10109.

Busecke, J.J.M., Spring, A., Maroon, E., Nicholas, T., Magin, J., Ritschel, M. & Angevaare, J.J.R. 2024. Available: https://github.com/jbusecke/xMIP [2025, January 16].

Cao, J. & Wang, B. 2019. DOI: 10.22033/ESGF/CMIP6.8769.

Cavalieri, D.J., Crawford, J.P., Drinkwater, M.R., Eppler, D.T., Farmer, L.D., Jentz, R.R. & Wackerman, C.C. 1991. Aircraft active and passive microwave validation of sea ice concentration from the Defense Meteorological Satellite Program special sensor microwave imager. *Journal of Geophysical Research*. 96:21989–22008.

Chai, Z. 2020. DOI: 10.22033/ESGF/CMIP6.3353.

Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., Ananthakrishnan, R., Miller, N., *et al*. 2014. The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*. 36:400–417. DOI: https://doi.org/10.1016/j.future.2013.07.002.

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.4271.

Durack, P.J., Taylor, K.E., Gleckler, P.J., Meehl, G.A., Lawrence, B.N., Covey, C., Stouffer, R.J., Levavasseur, G., *et al*. 2025. The Coupled Model Intercomparison Project (CMIP): Reviewing project history, evolution, infrastructure and implementation. EGUsphere. 2025:1–74. DOI: 10.5194/egusphere-2024-3729.

Global Circulation Model (GCM). n.d. Available: https://www.appsolutelydigital.com/ModelPrimer/chapter11_section2.html [2025, January 14].

Golden, K.M., Bennetts, L.G., Cherkaev, E., Eisenman, I., Feltham, D., Horvat, C., Hunke, E., Jones, C., *et al*. 2020. Modelling Sea Ice. Notices of the American Mathematical Society. 67(10):1. DOI: 10.1090/noti2171.

Goosse, H., Kay, J.E., Armour, K.C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., Jonko, A., Kushner, P.J., *et al*. 2018. Quantifying climate feedbacks in polar regions. Nature Communications. 9(1):1919. DOI: 10.1038/s41467-018-04173-0.

Guo, H., John, J.G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Rand, K., Zadeh, N.T., *et al*. 2018. DOI: 10.22033/ESGF/CMIP6.8594.

Haarsma, R.J., Roberts, M.J., Vidale, P.L., Senior, C.A., Bellucci, A., Bao, Q., Chang, P., Corti, S., *et al*. 2016. High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. Geoscientific Model Development. 9(11):4185–4208. DOI: 10.5194/gmd-9-4185-2016.

Hall, A. 2004. The Role of Surface Albedo Feedback in Climate. Journal of Climate. 17(7):1550–1568. DOI: 10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2.

Krasting, J.P., John, J.G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Rand, K., Zadeh, N.T., *et al*. 2018. DOI: 10.22033/ESGF/CMIP6.8597.

Lee, W.-L. & Liang, H.-C. 2020. DOI: 10.22033/ESGF/CMIP6.9755.

Li, L. 2019. DOI: 10.22033/ESGF/CMIP6.3356.

Lin, X., Massonnet, F., Fichefet, T. & Vancoppenolle, M. 2021. SITool (v1.0) – a new evaluation tool for large-scale sea ice simulations: application to CMIP6 OMIP. *Geoscientific Model Development*. 14(10):6331–6354. DOI: 10.5194/gmd-14-6331-2021.

Lovato, T. & Peano, D. 2020. DOI: 10.22033/ESGF/CMIP6.3825.

Lovato, T., Peano, D. & Butenschön, M. 2021. DOI: 10.22033/ESGF/CMIP6.13195.

McSweeney, R. 2018. Q&A: How do climate models work? Available: https://www.carbonbrief.org/qa-how-do-climate-models-work/ [2025, January 14].

Meehl, G.A. 1995. Global Coupled General Circulation Models. (June 1). DOI: 10.1175/1520-0477-76.6.951.

Meehl, G.A., Boer, G.J., Covey, C., Latif, M. & Stouffer, R.J. 1997. Intercomparison makes for a better climate model. Eos, Transactions American Geophysical Union. 78(41):445–451. DOI: 10.1029/97EO00276.

Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J. & Taylor, K.E. 2007. THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. Bulletin of the American Meteorological Society. 88(9):1383–1394. DOI: 10.1175/BAMS-88-9-1383.

Meehl, G.A., Goddard, L., Murphy, J., Stouffer, R.J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M.A., *et al*. 2009. Decadal Prediction: Can It Be Skilful? Bulletin of the American Meteorological Society. 90(10):1467–1486. DOI: 10.1175/2009BAMS2778.1.

Meehl, G.A., Moss, R., Taylor, K.E., Eyring, V., Stouffer, R.J., Bony, S. & Stevens, B. 2014. Climate Model Intercomparisons: Preparing for the Next Phase. Eos, Transactions American Geophysical Union. 95(9):77–78. DOI: 10.1002/2014EO090001.

Meier, W., Fetterer, F., Windnagel, A. & Stewart, S. 2021. NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4. DOI: 10.7265/EFMZ-2T65.

Nash, J.E. & Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*. 10(3):282–290. DOI: https://doi.org/10.1016/0022-1694(70)90255-6.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D.S., Tegen, I., Wieners, K.-H., Mauritsen, T., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.5016.

OSI SAF (2017): Global Sea Ice Concentration Climate Data Record v2.0 - Multimission, EUMETSAT SAF on Ocean and Sea Ice, DOI: 10.15770/EUM_SAF_OSI_0008, https://doi.org/10.15770/EUM_SAF_OSI_0008

Park, S. & Shin, J. 2019. DOI: 10.22033/ESGF/CMIP6.7789.

Roberts, M. 2017a. DOI: 10.22033/ESGF/CMIP6.6042.

Roberts, M. 2017b. DOI: 10.22033/ESGF/CMIP6.6045.

Roberts, M. 2018. DOI: 10.22033/ESGF/CMIP6.6040.

Rochford, P.A. 2016. SkillMetrics: A Python package for calculating the skill of model predictions against observations, http://github.com/PeterRochford/SkillMetrics

Rong, X. 2019. DOI: 10.22033/ESGF/CMIP6.9754.

Scoccimarro, E., Bellucci, A. & Peano, D. 2018. DOI: 10.22033/ESGF/CMIP6.3818.

Scoccimarro, E., Bellucci, A. & Peano, D. 2019. DOI: 10.22033/ESGF/CMIP6.3817.

Seland, Ø., Bentsen, M., Oliviè, D.J.L., Toniazzo, T., Gjermundsen, A., Graff, L.S., Debernard, J.B., Gupta, A.K., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.8036.

Selivanova, J., Iovino, D. & Vichi, M. (in press). Limited Benefits of Increased Spatial Resolution for Sea Ice in HighResMIP Simulations. *Geophysical Research Letters*. 51(14):e2023GL107969. DOI: https://doi.org/10.1029/2023GL107969.

Song, Z., Qiao, F., Bao, Y., Shu, Q., Song, Y. & Yang, X. 2019. DOI: 10.22033/ESGF/CMIP6.9199.

Swart, N.C., Cole, J.N.S., Kharin, V.V., Lazare, M., Scinocca, J.F., Gillett, N.P., Anstey, J., Arora, V., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.3610.

Tatebe, H. & Watanabe, M. 2018. DOI: 10.22033/ESGF/CMIP6.5603.

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Legutke, S., Schupfner, M., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.6595.

Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., Li, W., Liu, Q., *et al*. 2018. DOI: 10.22033/ESGF/CMIP6.2948.

YU, Y. 2019. DOI: 10.22033/ESGF/CMIP6.3355.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.6842.

Zhang, J., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., Liu, Q., Yan, J., *et al*. 2018. DOI: 10.22033/ESGF/CMIP6.2949.

Zhao, M., Blanton, C., John, J.G., Radhakrishnan, A., Zadeh, N.T., McHugh, C., Rand, K., Vahlenkamp, H., *et al*. 2018. DOI: 10.22033/ESGF/CMIP6.8567.

Zhuang, J., Dussin, Raphael, Huard, D., Bourgault, P., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., *et al*. 2024. DOI: 10.5281/zenodo.14025505.

Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., Mackallah, C., Druken, K., *et al*. 2019. DOI: 10.22033/ESGF/CMIP6.4232.

Appendix A
HighResMIP Additional Metric Plots

A.1 GFDL-CM4C192

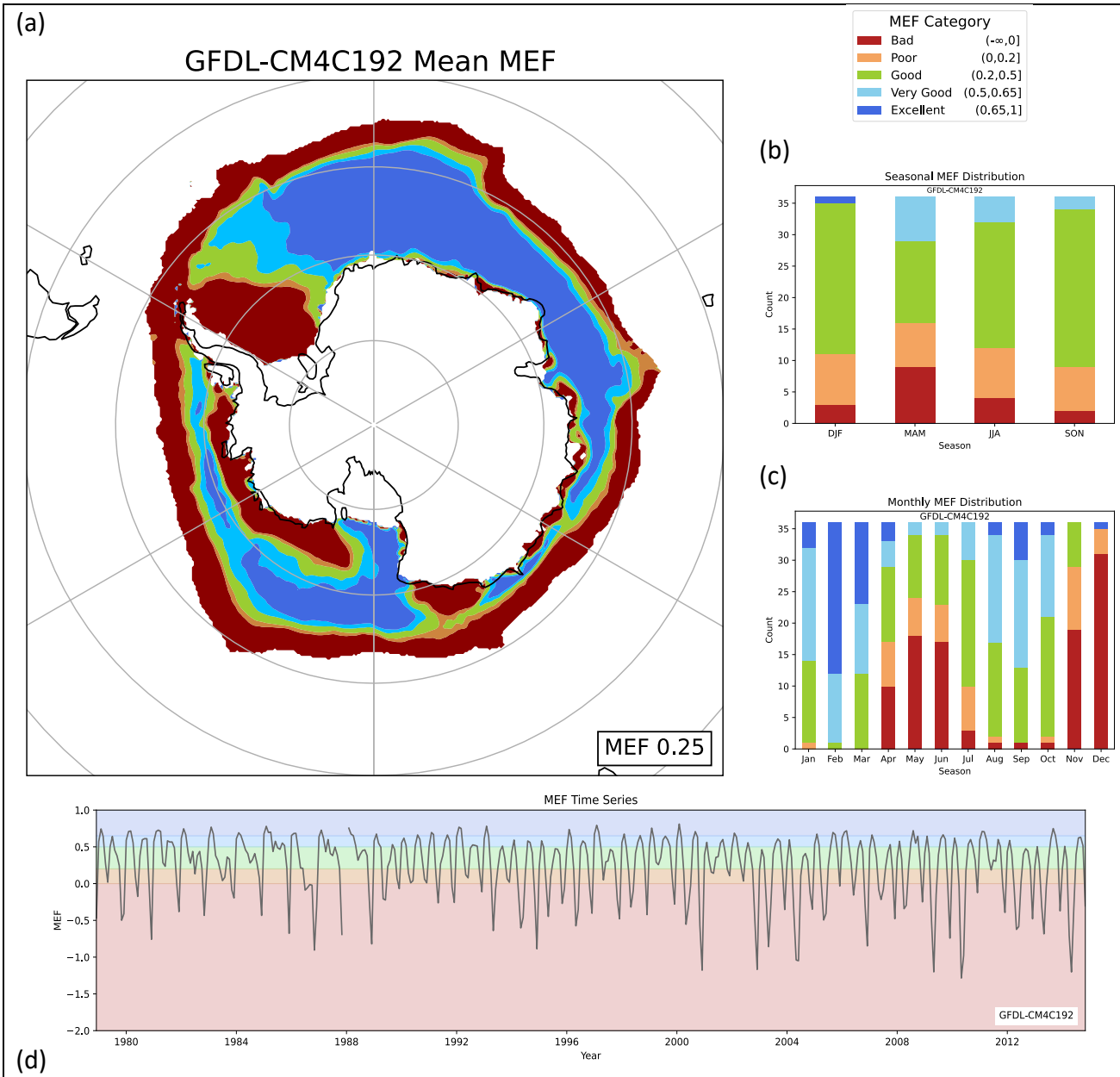


Figure A-1 GFDL-CM4C192 MEF Plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

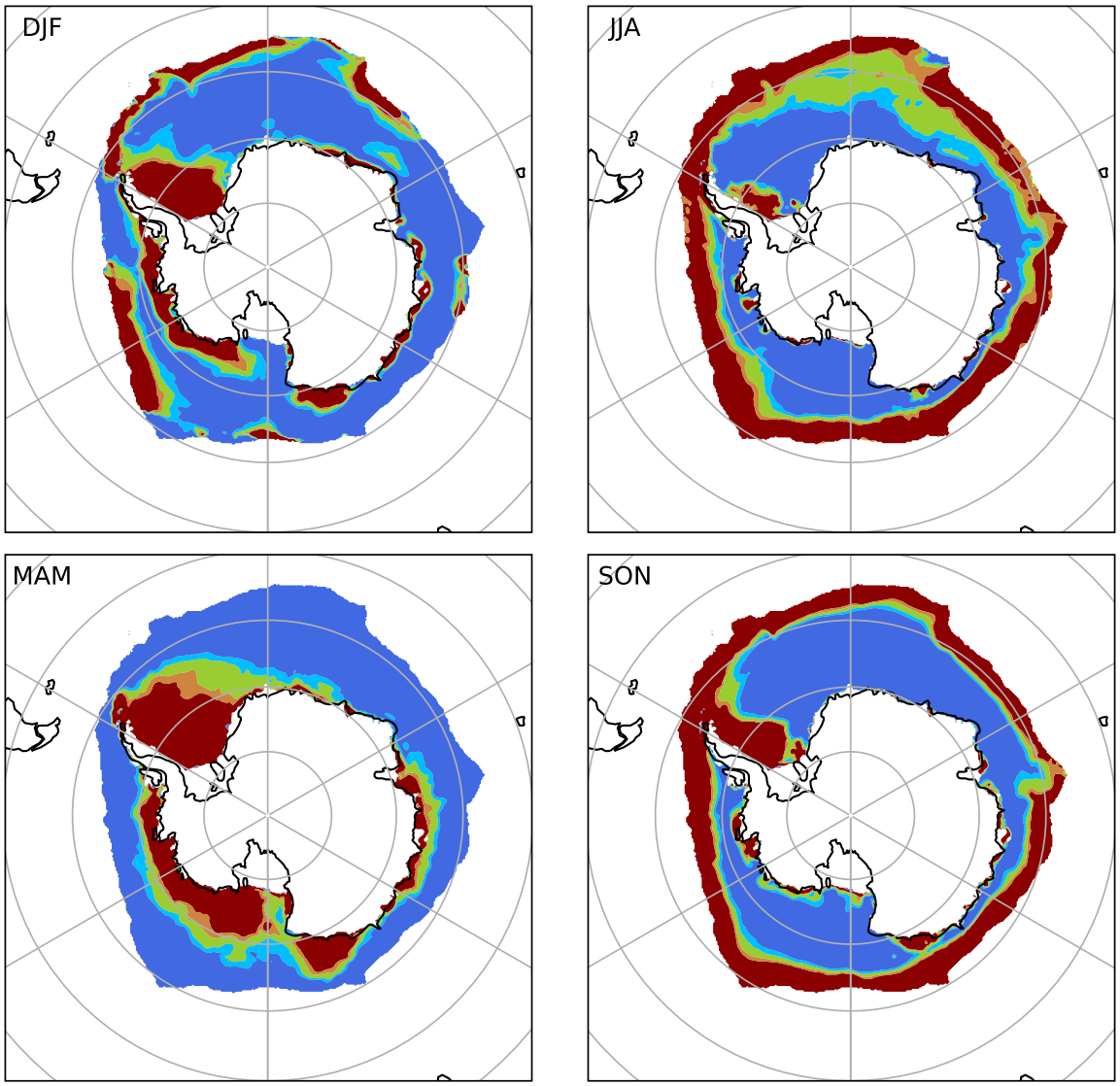


Figure A-2 GFDL-CM4C192 Seasonal Spatial MEF Plots.
Values colour mapped to categorical performance colours defined in Figure 3-1.

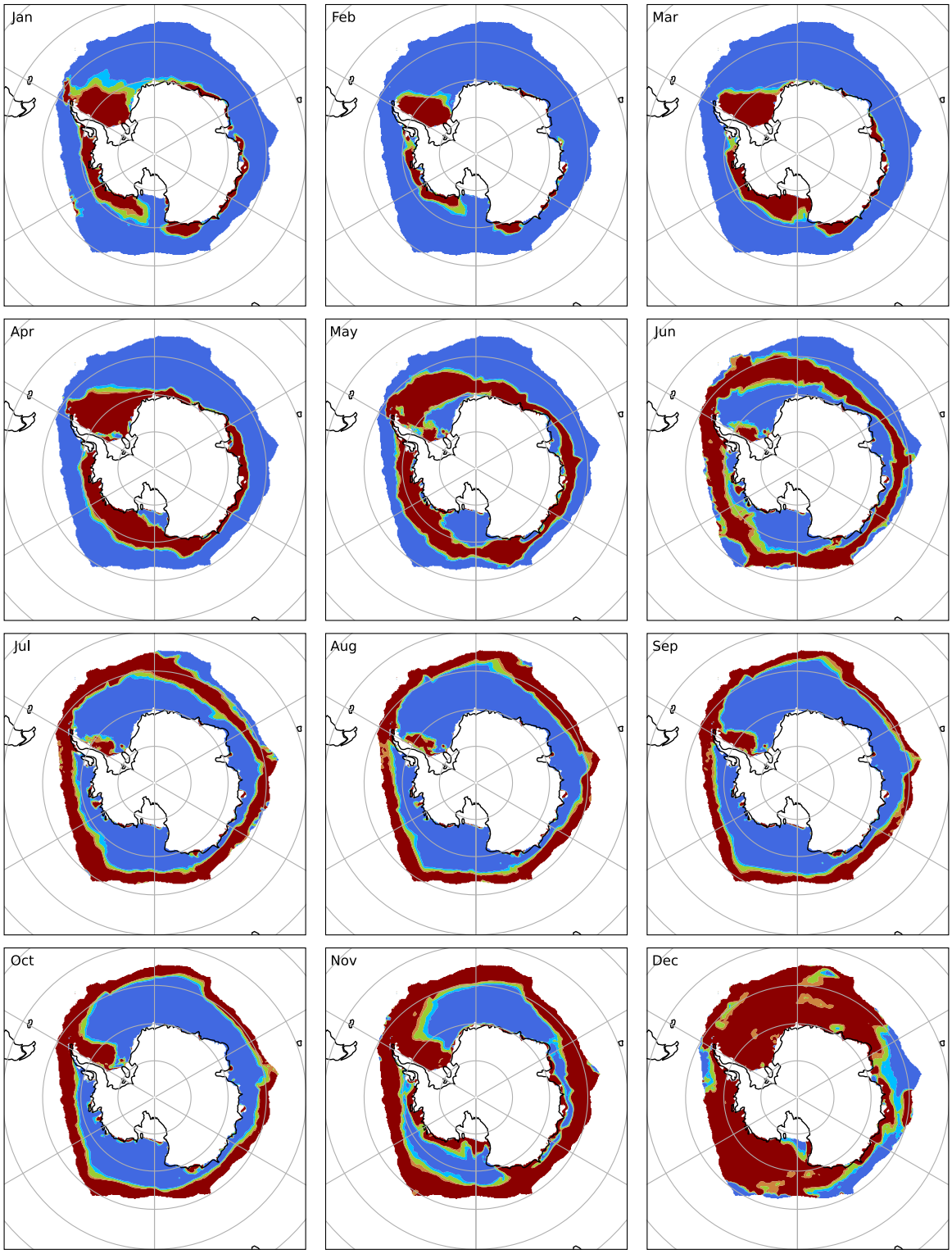
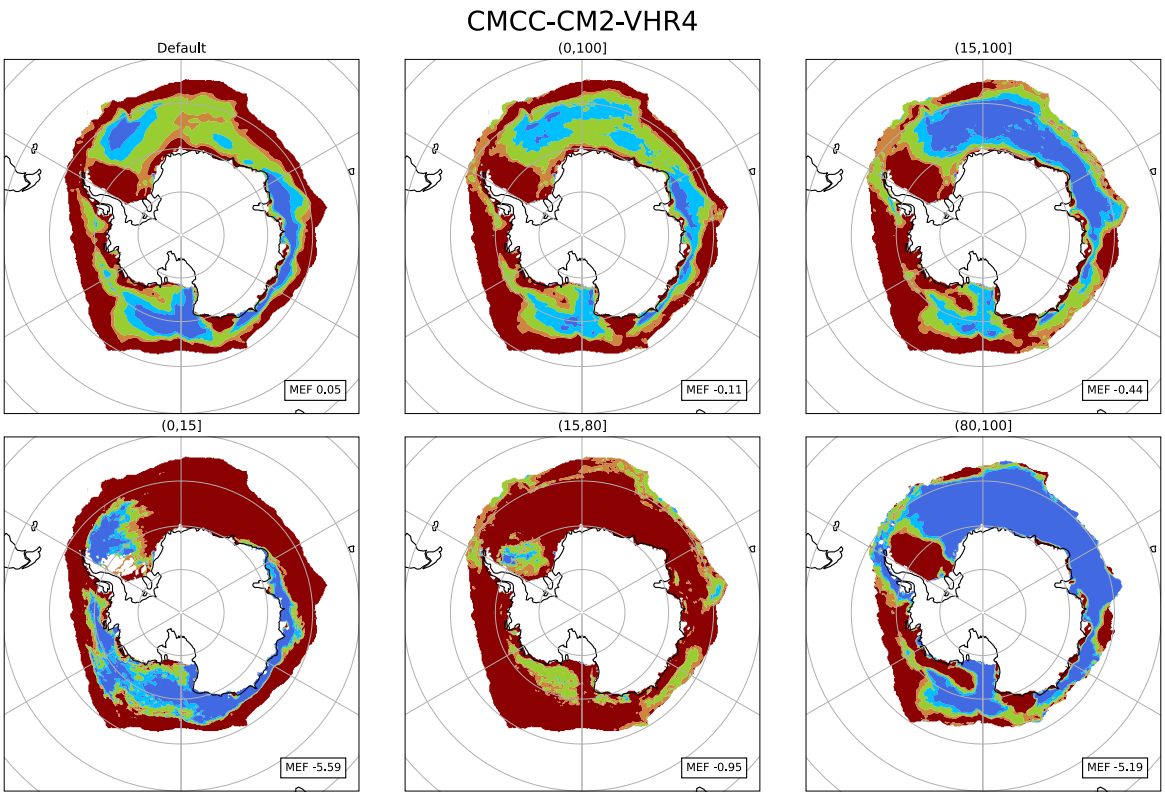


Figure A-3 GFDL-CM4C192 Monthly Spatial MEF Plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

A.2 Omitted Sensitivity Analysis Plots

(a)



(b)

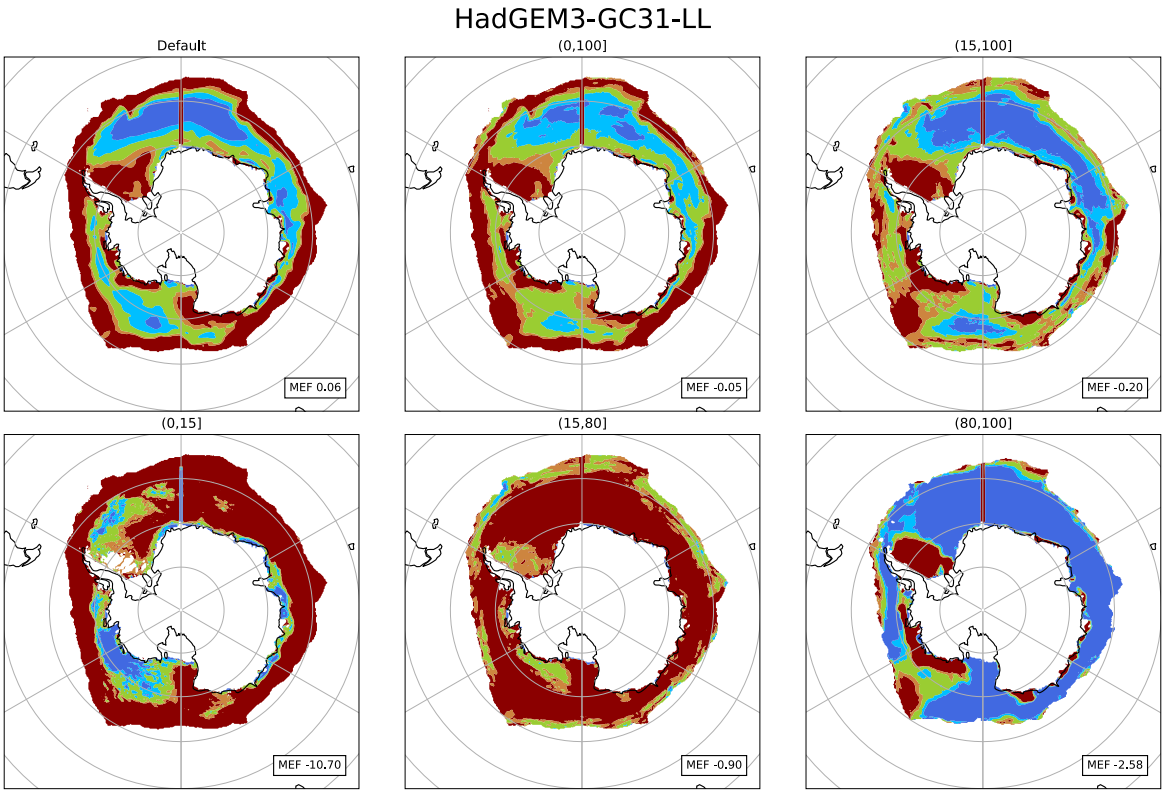
CMCC-CM2-VHR4

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.05	-0.11	0.01	0.03	0.28	0.20	0.61	0.59	0.15	-0.71	-0.49	0.15	0.42	0.51	0.46	-0.13	-1.15
(0,100]	-0.11	-0.53	-0.54	0.25	0.36	-0.12	-0.10	-0.17	-0.60	-0.85	-0.17	0.36	0.55	0.58	0.51	-0.02	-1.37
(15,100]	-0.44	-1.45	-1.15	0.40	0.43	-1.10	-1.44	-1.25	-1.40	-0.81	0.08	0.49	0.62	0.63	0.56	0.11	-1.81
(0,15]	-5.59	0.11	-0.08	-10.30	-11.96	0.46	0.64	0.56	0.16	-0.97	-3.05	-8.76	-19.09	-20.57	-10.06	-5.25	-0.79
(15,80]	-0.95	-0.64	-1.74	-1.10	-0.31	-0.07	-0.04	-0.35	-1.40	-3.46	-2.36	-0.78	-0.15	0.01	0.02	-0.96	-1.81
(80,100]	-5.19	-14.50	-7.81	0.72	0.64	-12.81	-28.29	-22.06	-1.54	0.16	0.59	0.75	0.80	0.79	0.71	0.42	-1.99

Figure A-4 CMCC-CM2-VHR4 Variation Plots

Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat mapped values follow categories defined in Figure 3-1.

(a)



(b)

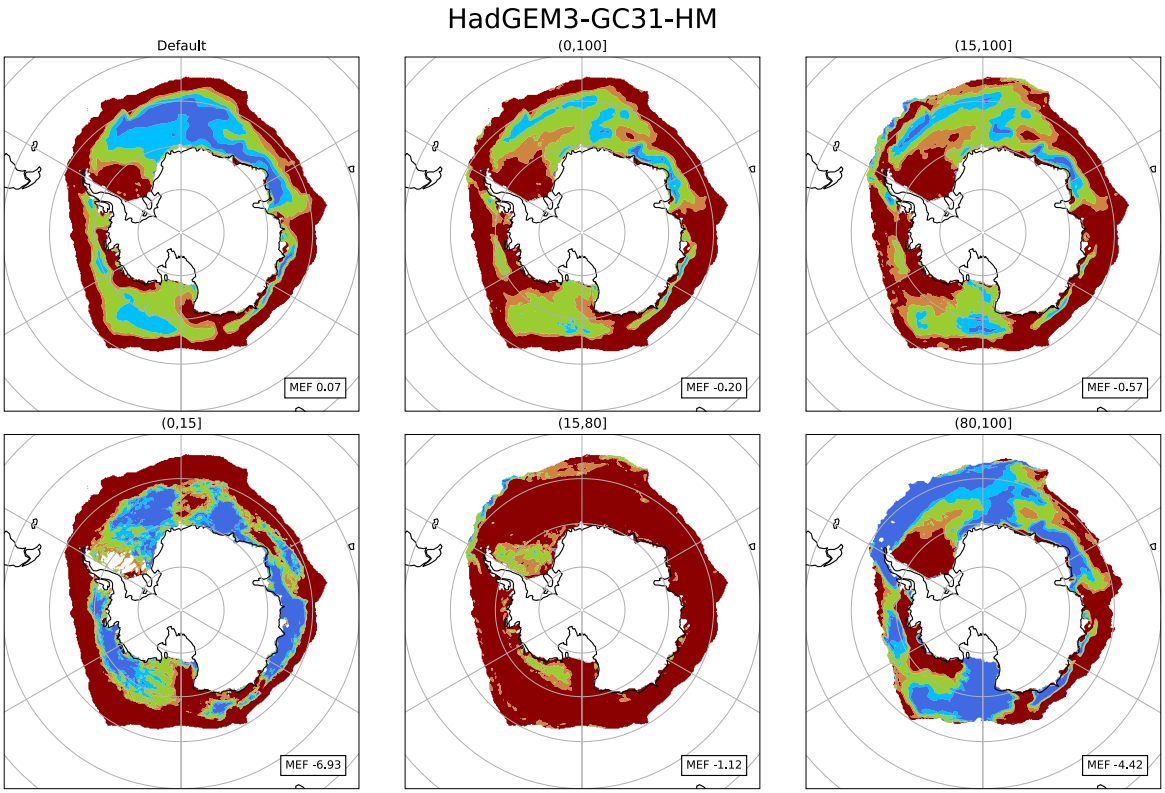
HadGEM3-GC31-LL

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.06	-0.21	0.19	0.13	0.15	0.26	0.64	0.54	0.25	-0.22	-0.16	0.17	0.36	0.44	0.38	-0.37	-1.55
(0,100]	-0.05	-0.38	-0.51	0.27	0.41	-0.02	0.12	-0.17	-0.65	-0.72	-0.04	0.35	0.51	0.57	0.55	0.10	-1.26
(15,100]	-0.20	-0.78	-1.04	0.42	0.59	-0.45	-0.74	-0.98	-1.37	-0.78	0.14	0.49	0.63	0.68	0.68	0.39	-1.14
(0,15]	-10.70	-0.22	-0.04	-14.50	-27.82	0.23	0.57	0.36	0.05	-0.54	-2.15	-9.94	-31.42	-43.34	-27.96	-12.17	-1.48
(15,80]	-0.90	-0.59	-1.77	-0.94	-0.29	-0.01	0.07	-0.56	-1.53	-3.22	-2.01	-0.65	-0.14	0.03	0.06	-0.97	-1.86
(80,100]	-2.58	-7.48	-4.46	0.71	0.83	-5.74	-16.54	-12.18	-1.32	0.11	0.60	0.72	0.81	0.85	0.86	0.79	0.10

Figure A-5 HadGEM3-GC31-LL Sensitivity Analysis Plots

Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat mapped values follow categories defined in Figure 3-1.

(a)



(b)

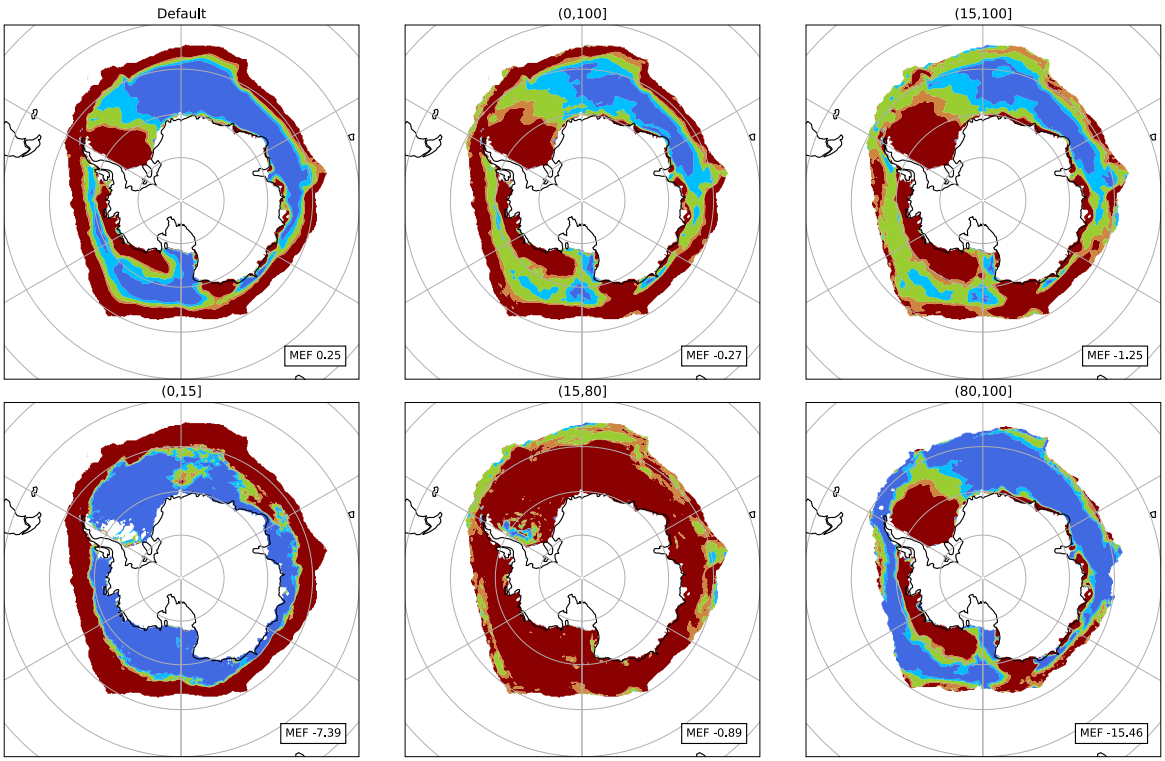
HadGEM3-GC31-HM

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.07	0.11	0.20	-0.23	0.21	0.45	0.70	0.64	0.35	-0.39	-0.68	-0.19	0.19	0.35	0.32	-0.04	-0.84
(0,100]	-0.20	-0.24	-0.76	-0.17	0.38	0.06	0.03	-0.21	-0.80	-1.28	-0.73	-0.07	0.28	0.43	0.46	0.24	-0.84
(15,100]	-0.57	-0.88	-1.78	-0.12	0.50	-0.58	-1.12	-1.35	-2.14	-1.85	-0.72	0.01	0.36	0.51	0.56	0.43	-0.92
(0,15]	-6.93	0.12	0.44	-8.94	-19.22	0.45	0.65	0.56	0.52	0.25	-0.89	-6.55	-19.38	-29.29	-20.93	-7.43	-0.75
(15,80]	-1.12	-0.43	-1.96	-1.60	-0.49	0.03	0.03	-0.53	-1.59	-3.77	-2.97	-1.28	-0.55	-0.29	-0.29	-0.88	-1.36
(80,100]	-4.42	-10.22	-8.56	0.20	0.78	-7.59	-22.55	-21.18	-3.35	-1.16	-0.24	0.28	0.57	0.72	0.80	0.82	-0.15

Figure A-6 HadGEM3-GC31-HM Sensitivity Analysis Plots

Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat mapped values follow categories defined in Figure 3-1.

(a)



(b)

GFDL-CM4C192

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.25	0.26	0.19	0.25	0.30	0.50	0.67	0.56	0.20	-0.19	-0.02	0.31	0.48	0.53	0.43	-0.07	-0.40
(0,100]	-0.27	-0.66	-1.24	0.30	0.50	-0.42	-0.90	-1.00	-1.59	-1.13	-0.11	0.41	0.60	0.65	0.60	0.23	-0.65
(15,100]	-1.25	-2.74	-3.28	0.36	0.64	-2.62	-4.34	-3.91	-4.12	-1.81	-0.10	0.49	0.68	0.75	0.72	0.44	-1.22
(0,15]	-7.39	0.66	0.87	-8.54	-22.40	0.90	0.97	0.97	0.91	0.72	-0.07	-4.93	-20.62	-35.03	-23.93	-8.23	0.09
(15,80]	-0.89	-0.68	-1.88	-0.85	-0.15	-0.45	-0.67	-0.92	-1.67	-3.06	-1.84	-0.63	-0.09	0.09	0.07	-0.62	-0.91
(80,100]	-15.46	-34.83	-28.82	0.61	0.85	-27.23	-74.13	-75.22	-9.84	-1.40	0.26	0.72	0.86	0.92	0.90	0.75	-2.01

Figure A-7 GFDL-CM4C192 Sensitivity Analysis Plots

Cumulative annual temporal mean MEF (a) for the default configuration and all five sensitivity analysis configurations. Aggregated mean MEF annual, seasonal, and monthly values (b) for all configurations. Colour contours and heat mapped values follow categories defined in Figure 3-1.

A.3 Sensitivity Analysis Spatial Mean Distributions

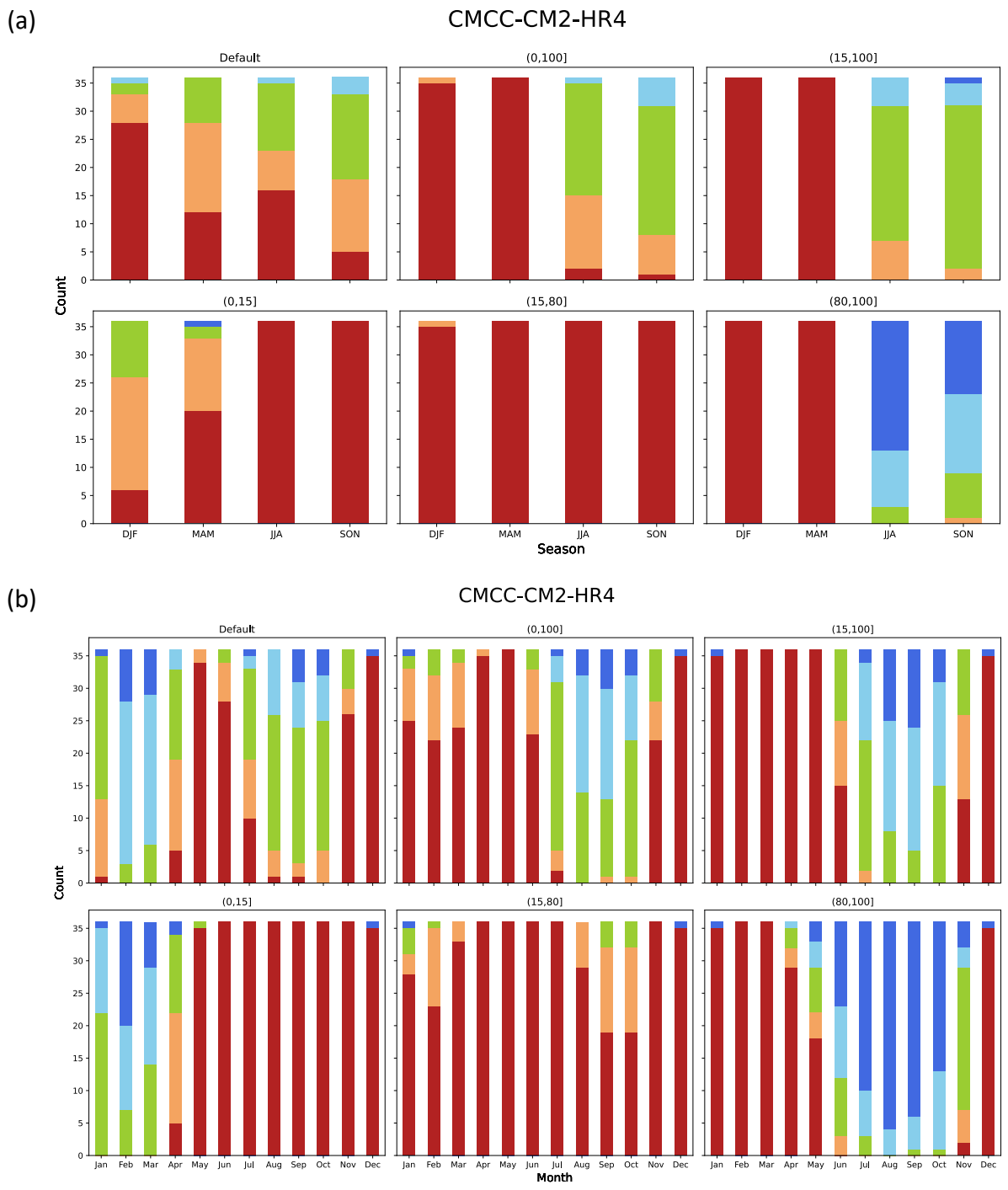


Figure A-8 CMCC-CM2-HR4 Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

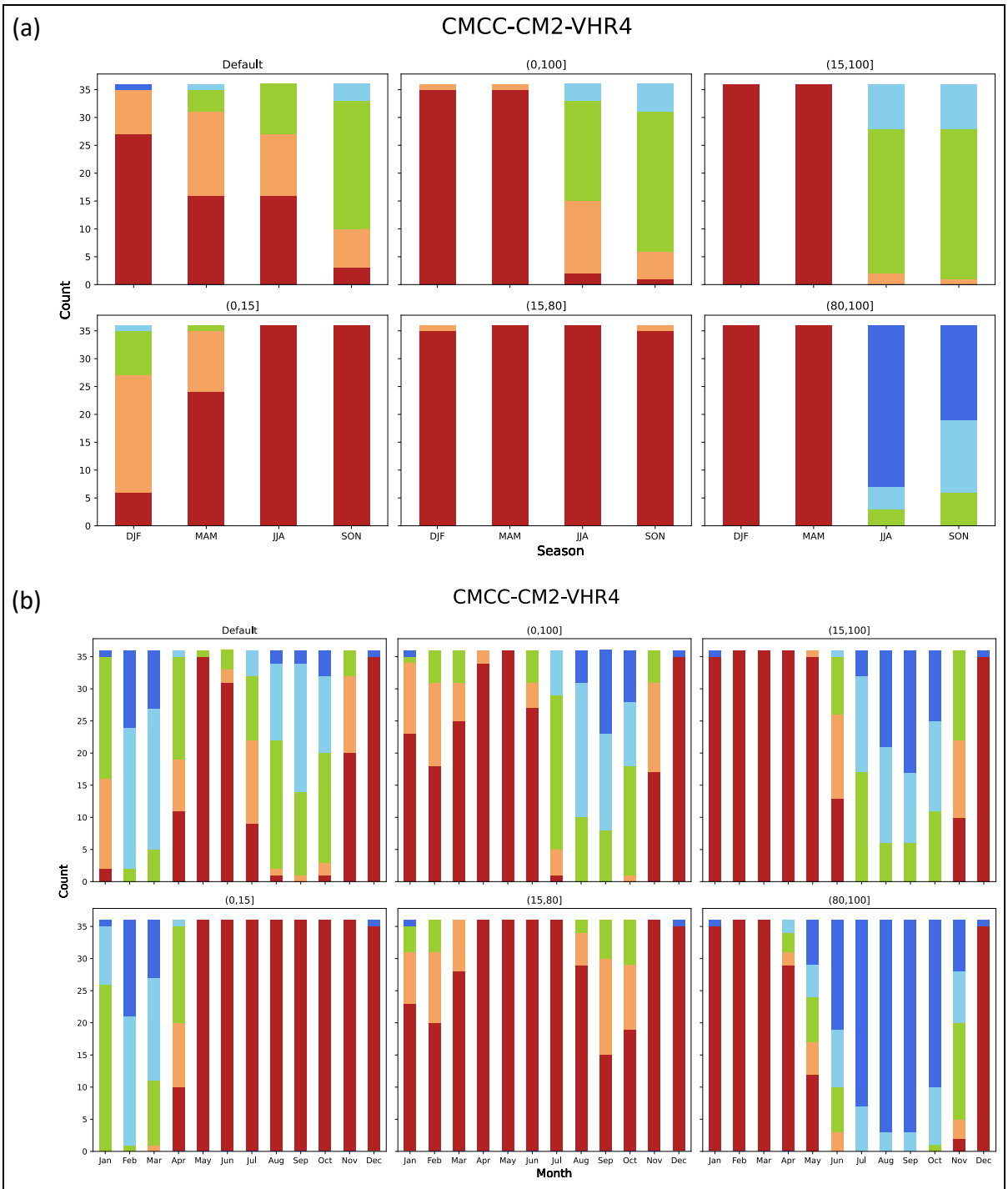


Figure A-9 CMCC-CM2-VHR4 Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

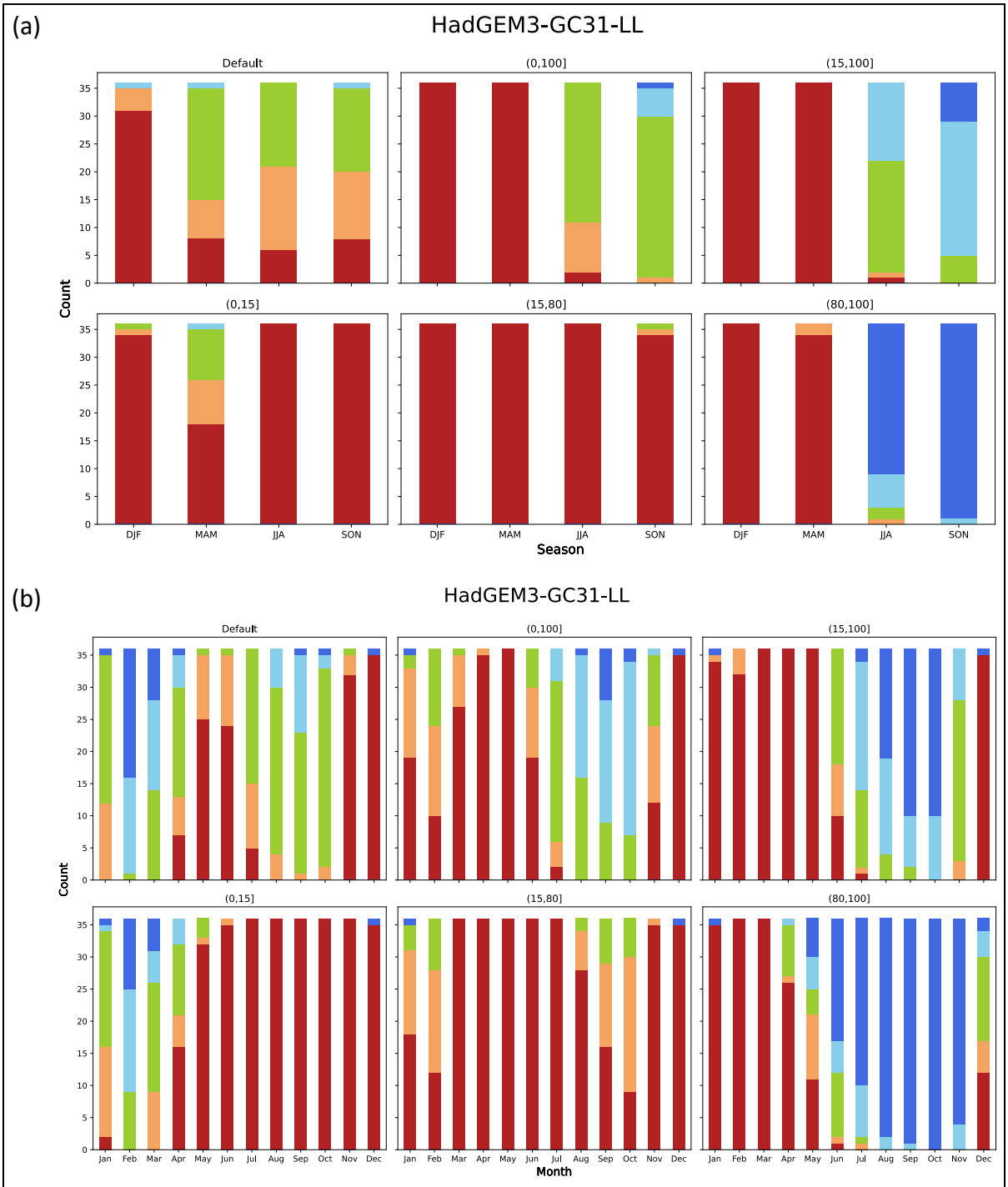


Figure A-10 HadGEM3-GC31-LL Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

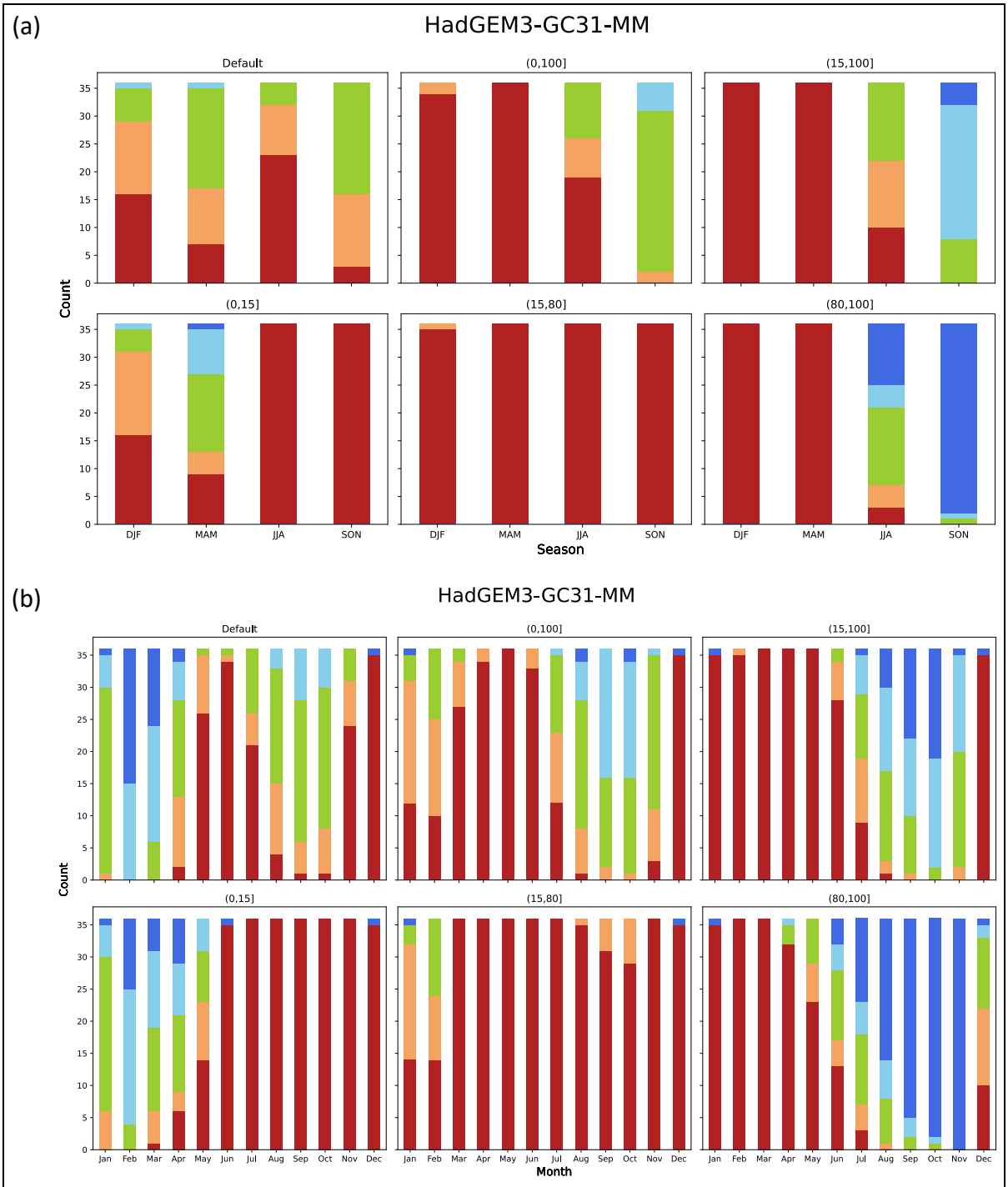


Figure A-11 HadGEM3-GC31-MM Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

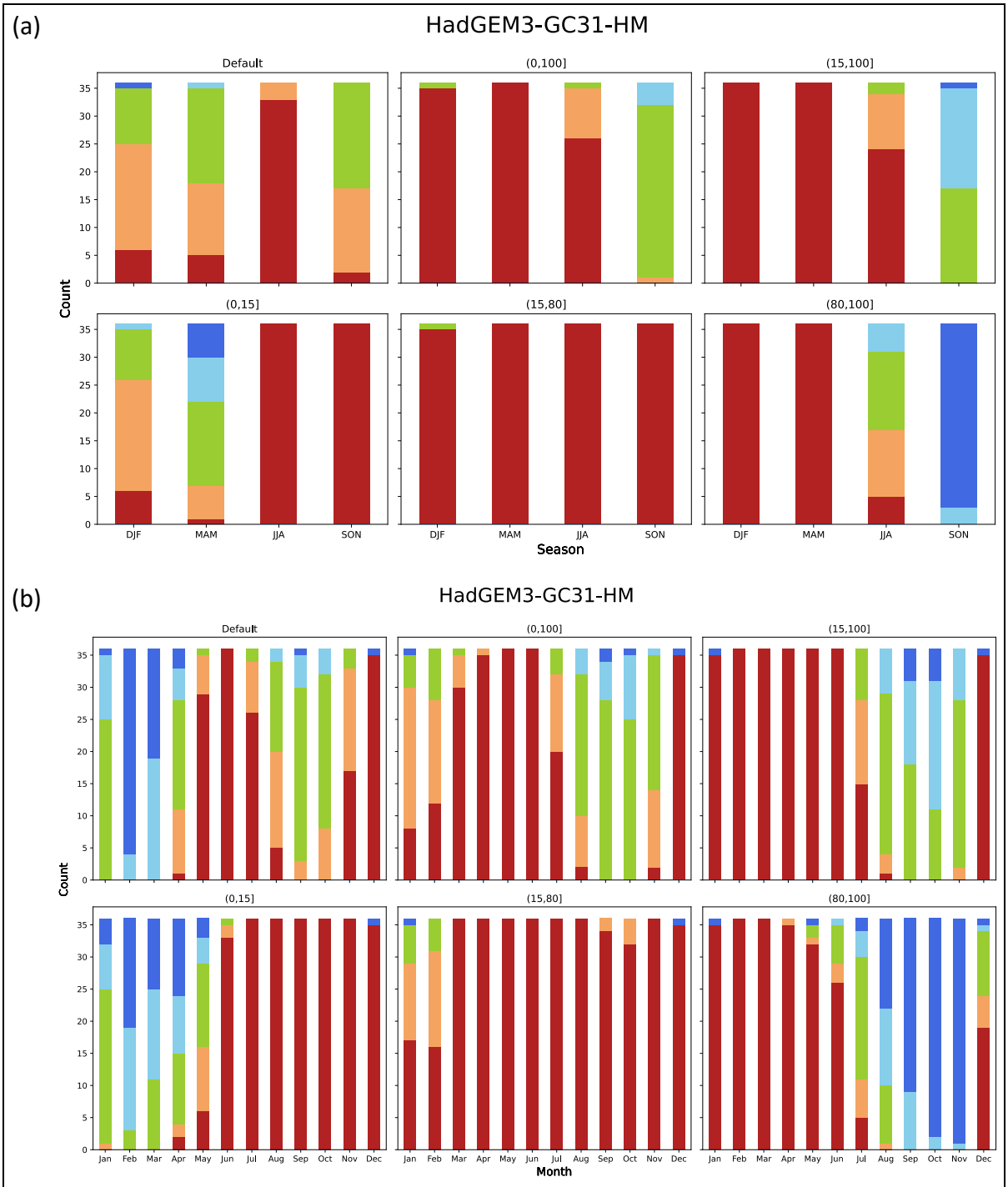


Figure A-12 HadGEM3-GC31-HM Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

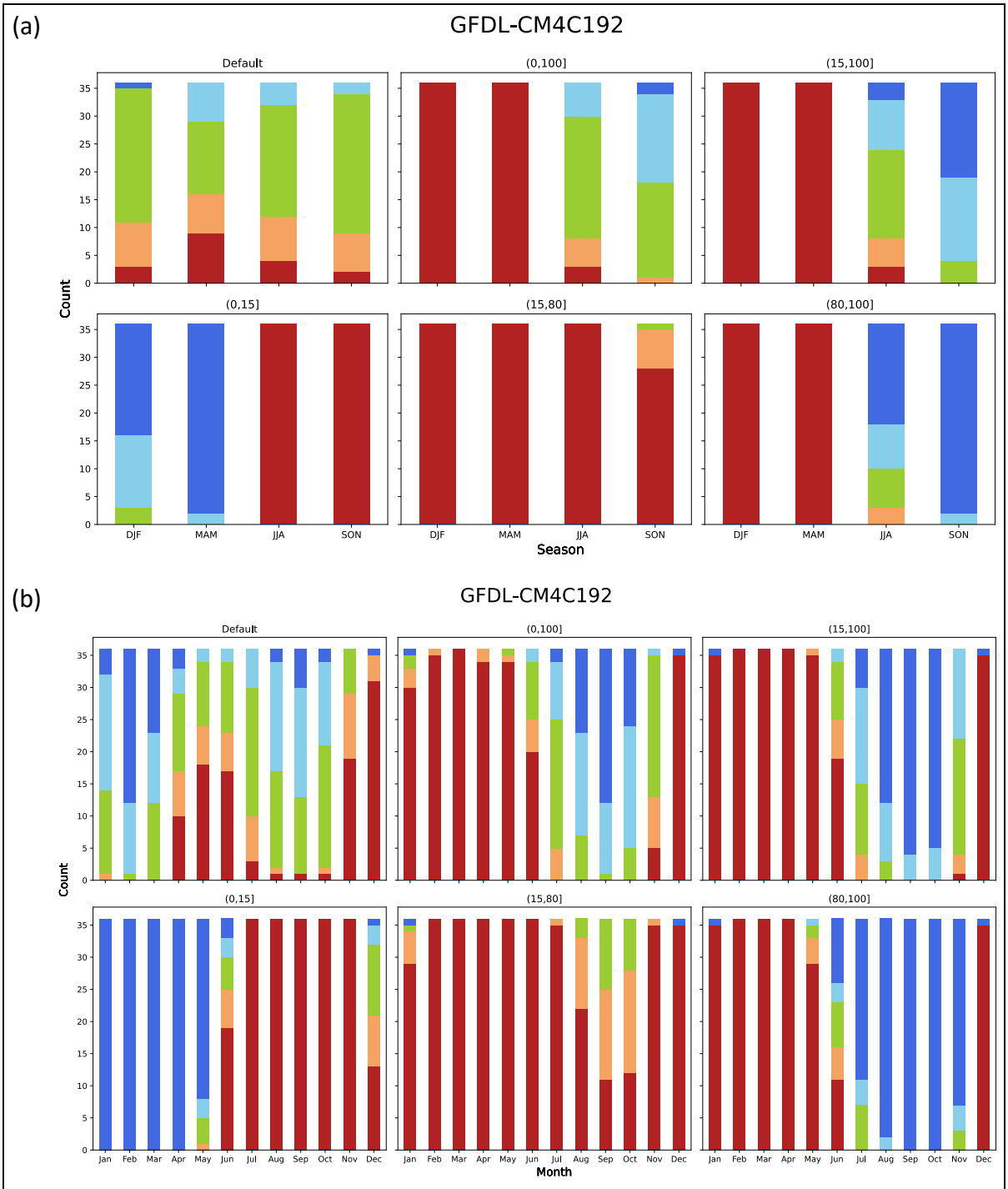


Figure A-13 GFDL-CM4C192 Sensitivity Analysis Distributions

(a) Seasonal temporal mean MEF distribution. (b) Monthly temporal mean MEF distribution. Colour contours and heat mapped values follow categories defined in Figure 3-1

Appendix B
CMIP6 Models Metric Plots

This appendix provides the plots produced from the initial toolset testing of the 24 CMIP6 models listed in Table 1. The provided plots include the following; (a) Cumulative temporal mean spatial plot, (b) Spatial mean seasonal distribution, (c) Spatial mean monthly distribution, (d) Spatial mean time series, and (e) Aggregated annual, seasonal and monthly values. All plots in this Appendix follow the colour categorisation of Figure 3-1.

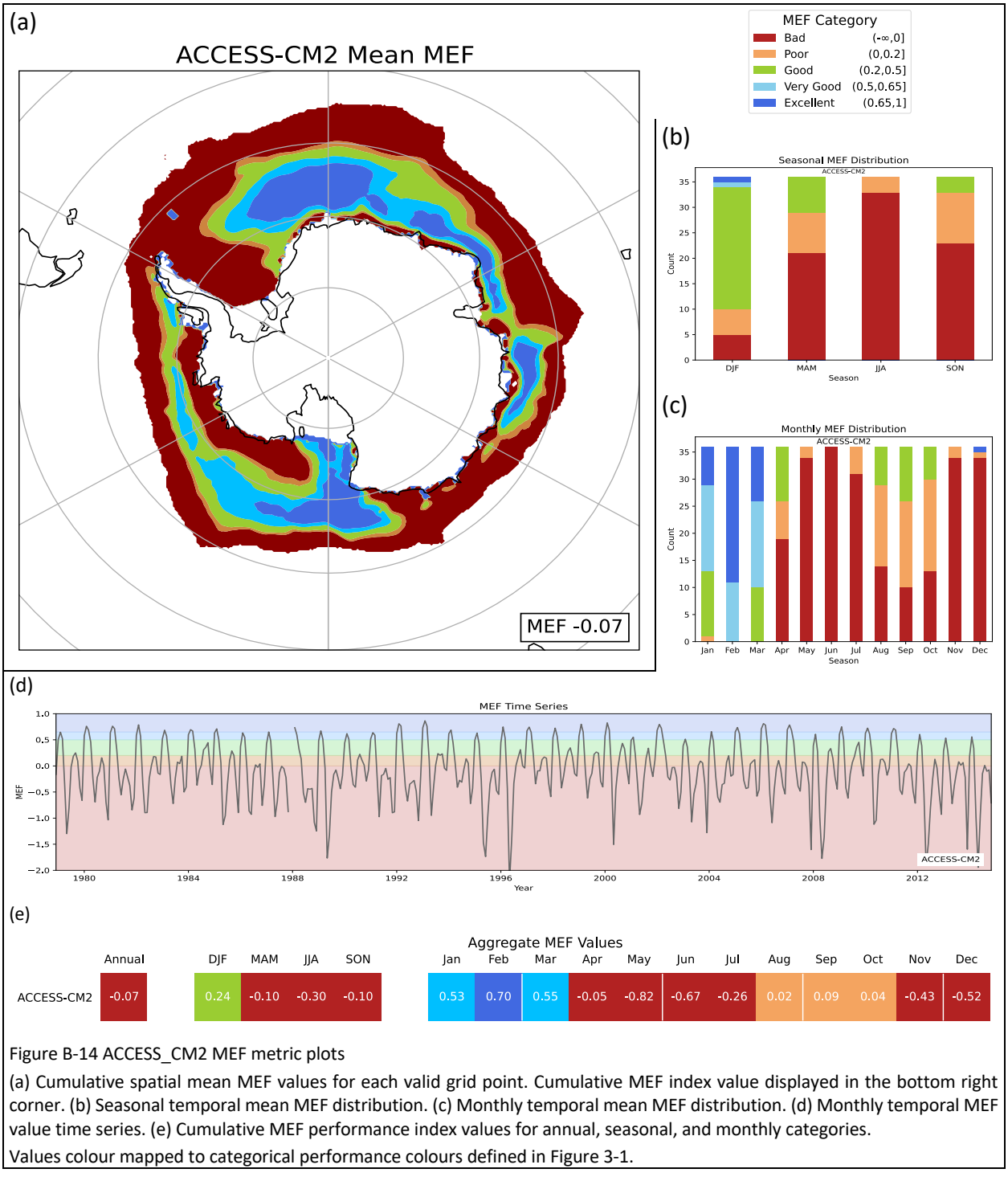


Figure B-14 ACCESS_CM2 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

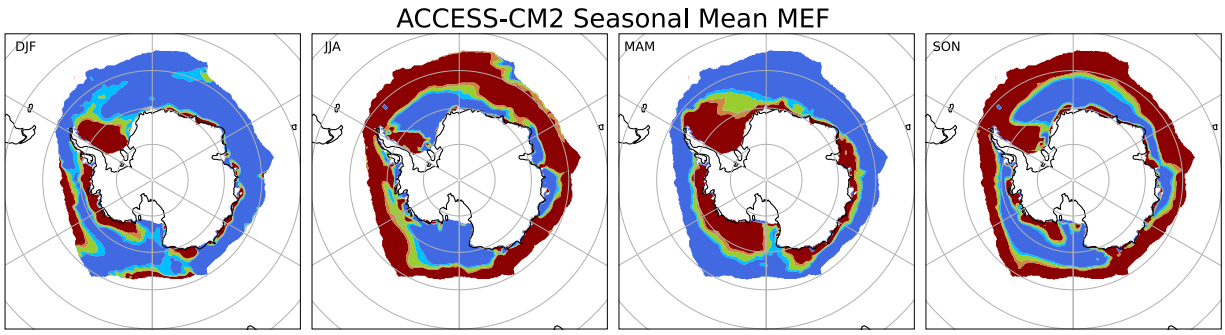


Figure B-15 ACCESS_CM2 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

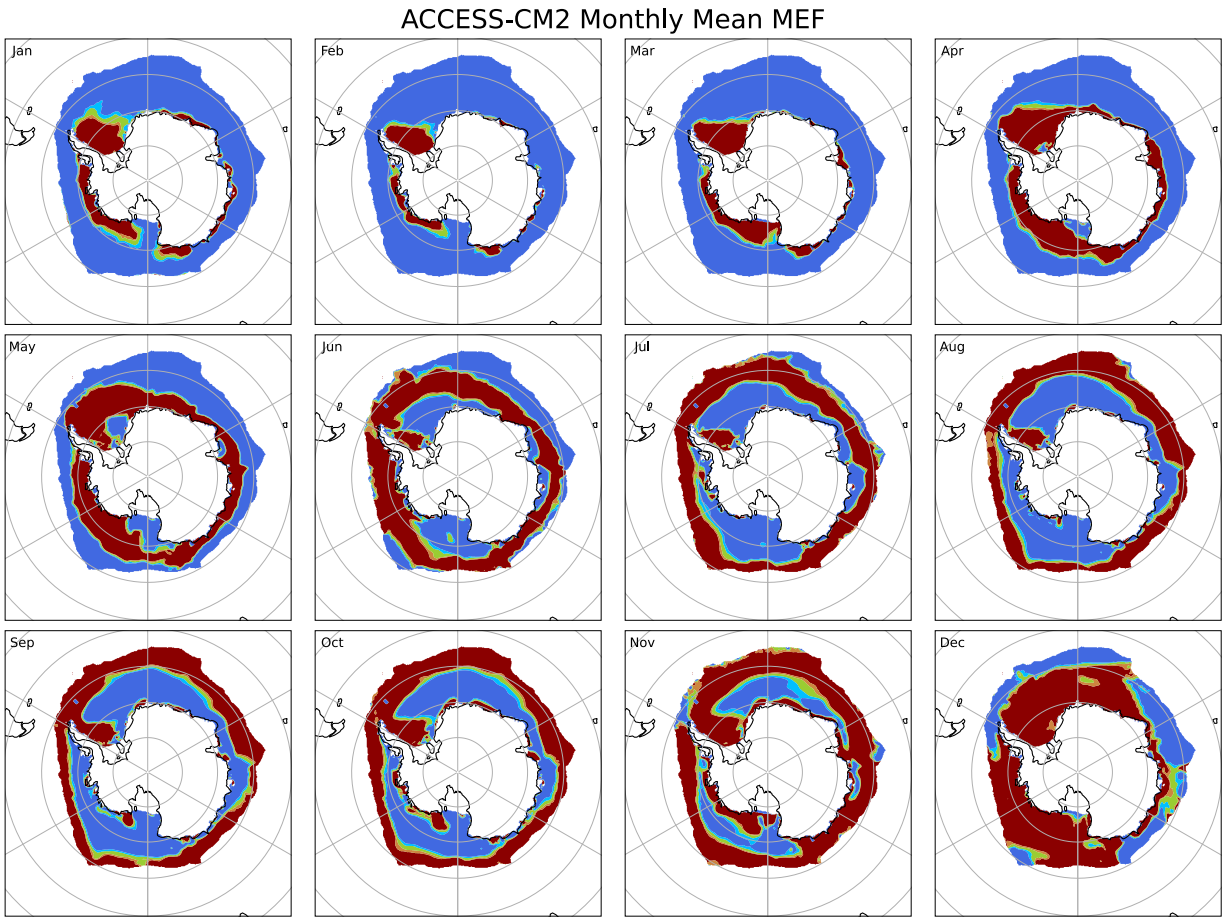


Figure B-16 ACCESS_CM2 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

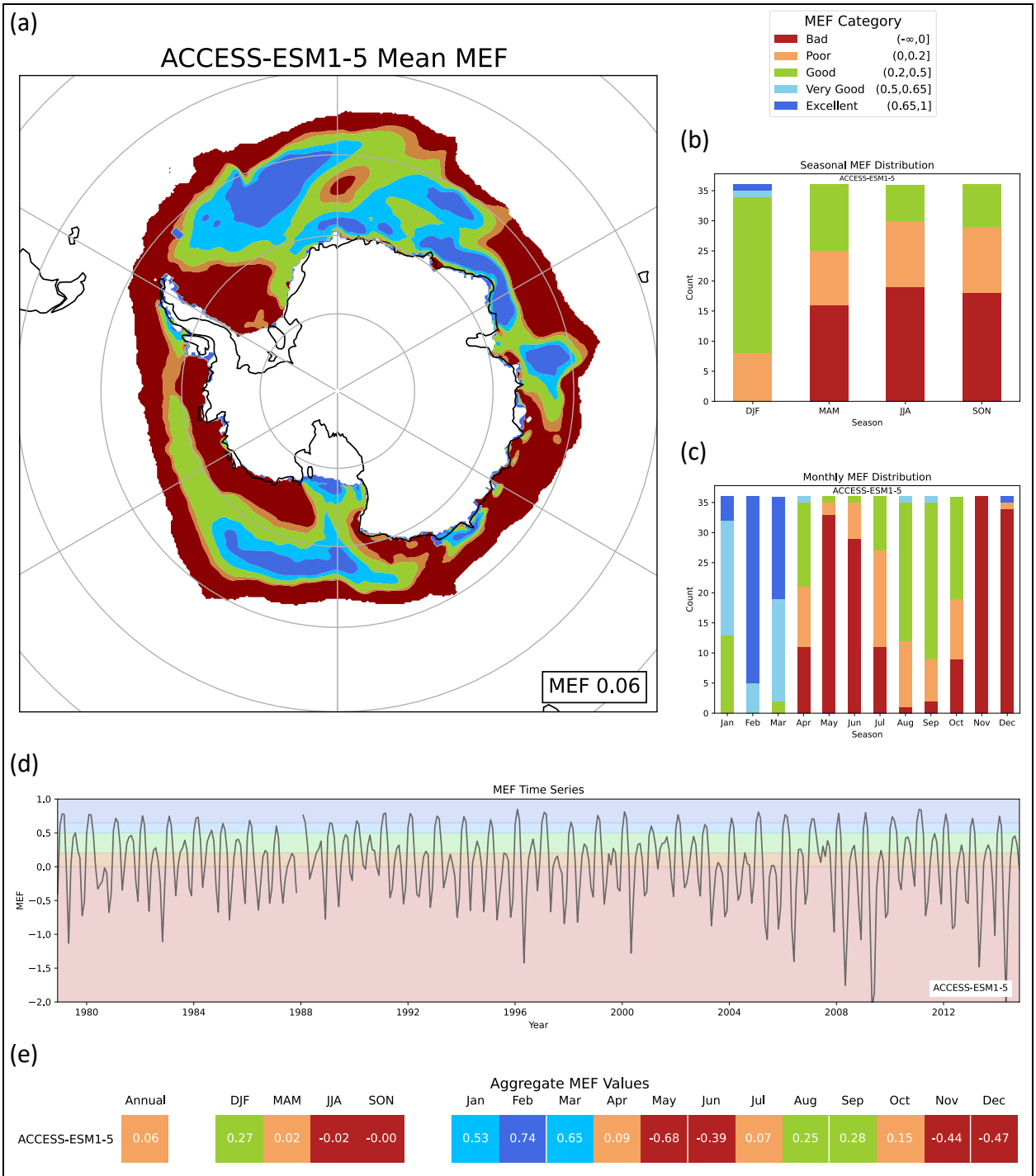


Figure B-17 ACCESS-ESM1-5 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

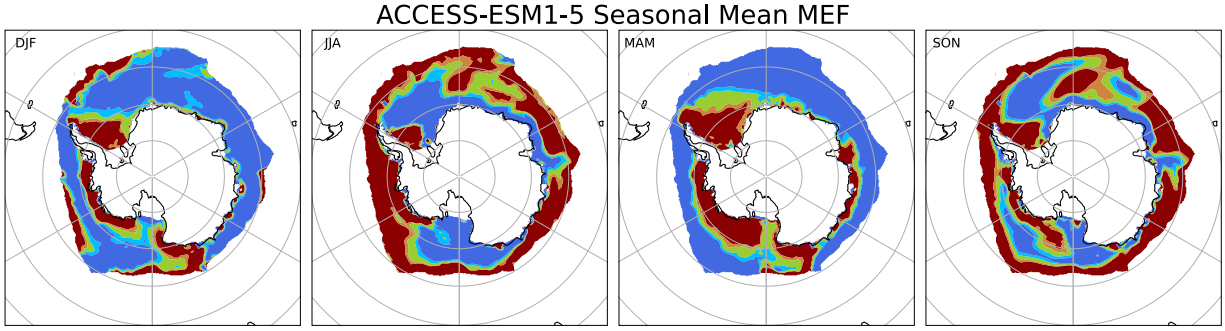


Figure B-18 ACCESS-ESM1-5 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

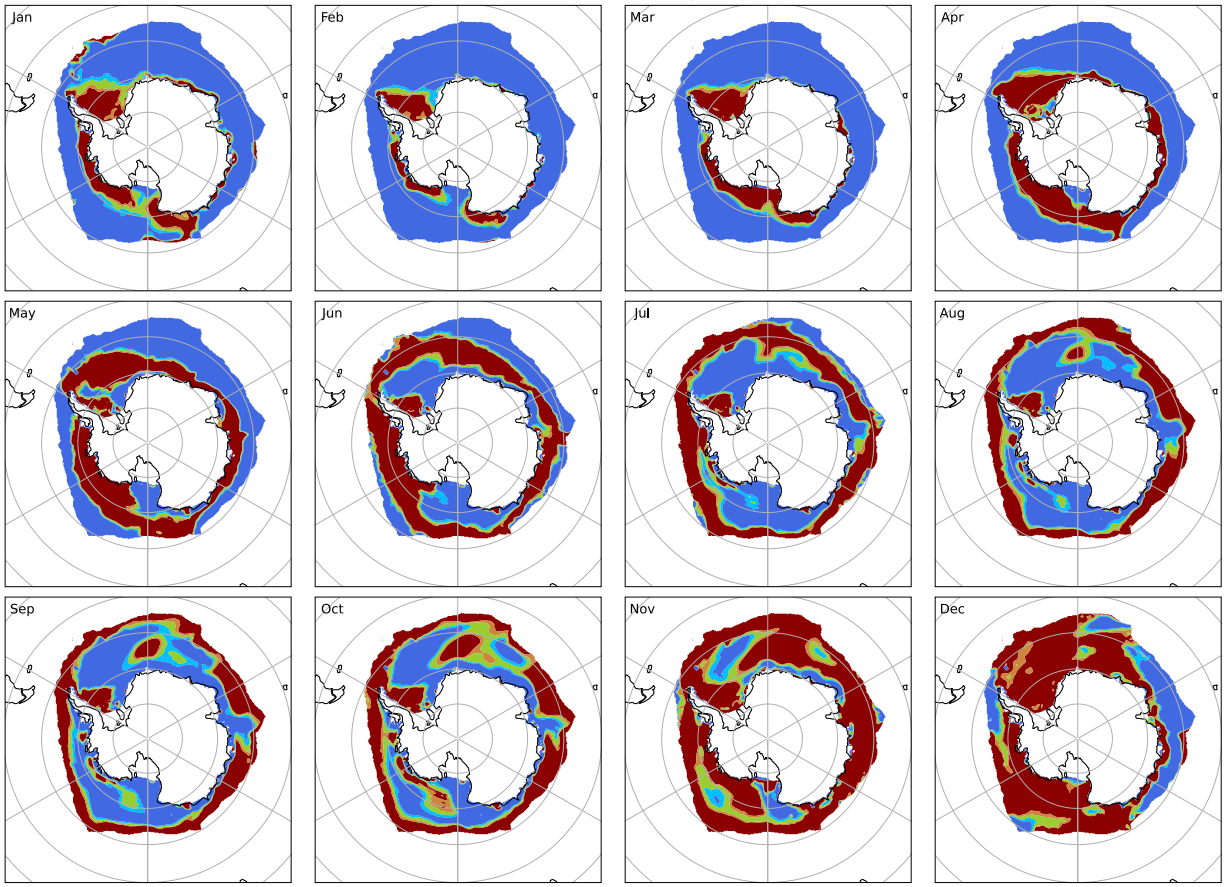


Figure B-19 ACCESS-ESM1-5 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

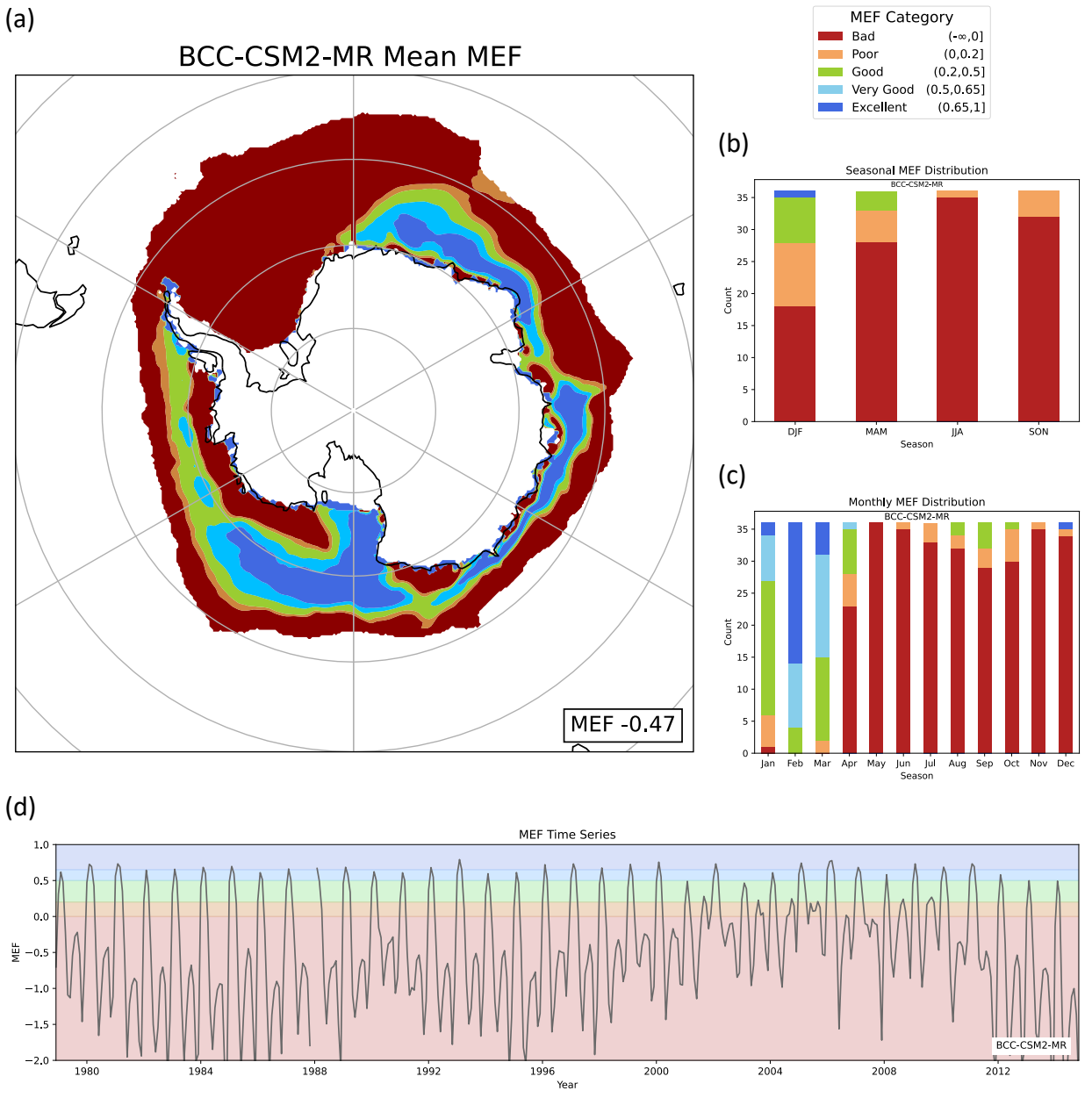


Figure B-20 BCC-CSM2-MR MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

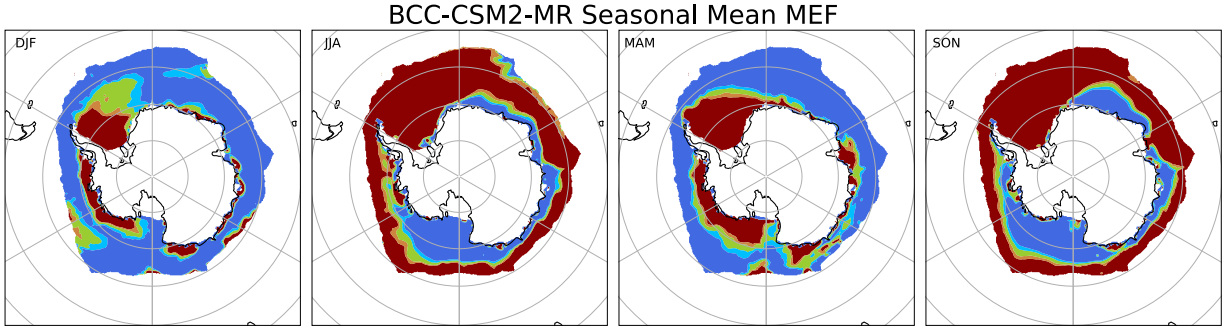


Figure B-21 BCC-CSM2-MR seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

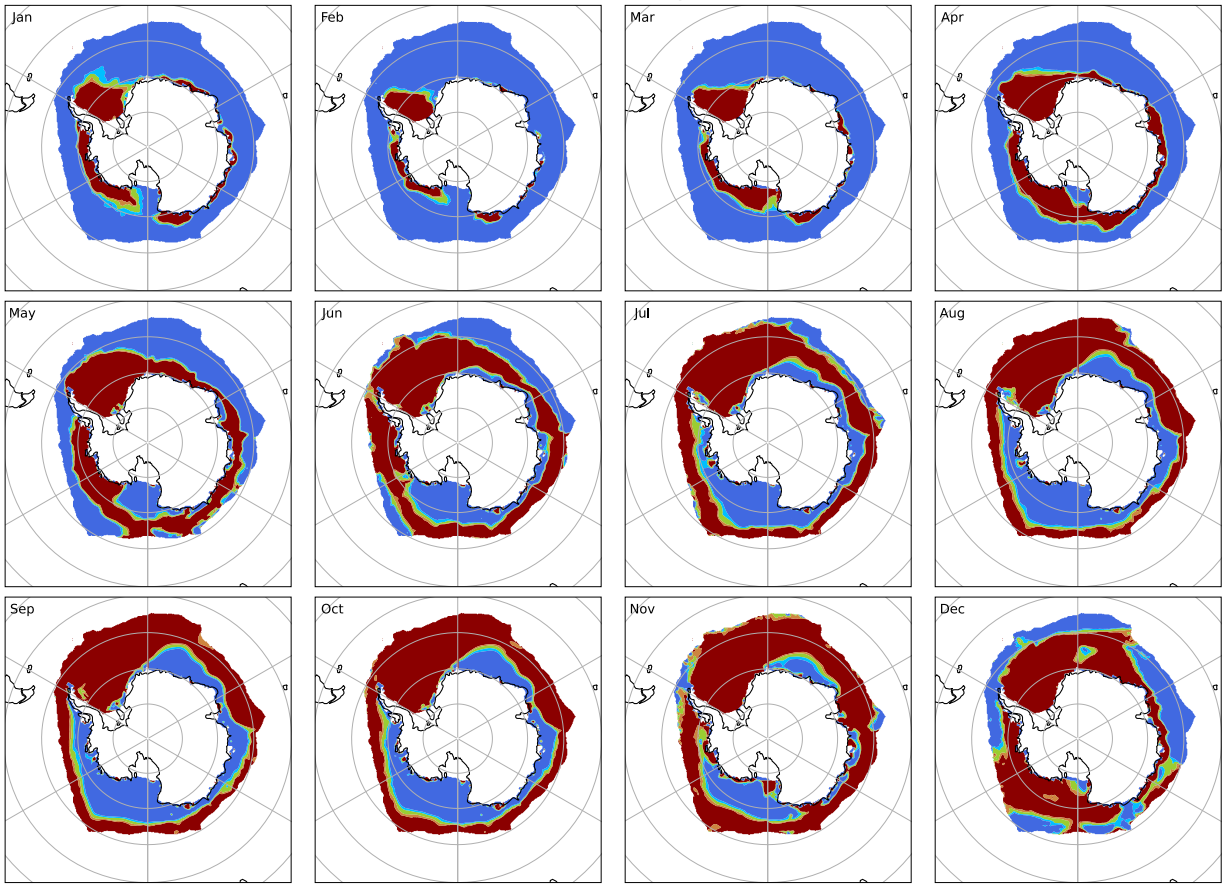


Figure B-22 BCC-CSM2-MR monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

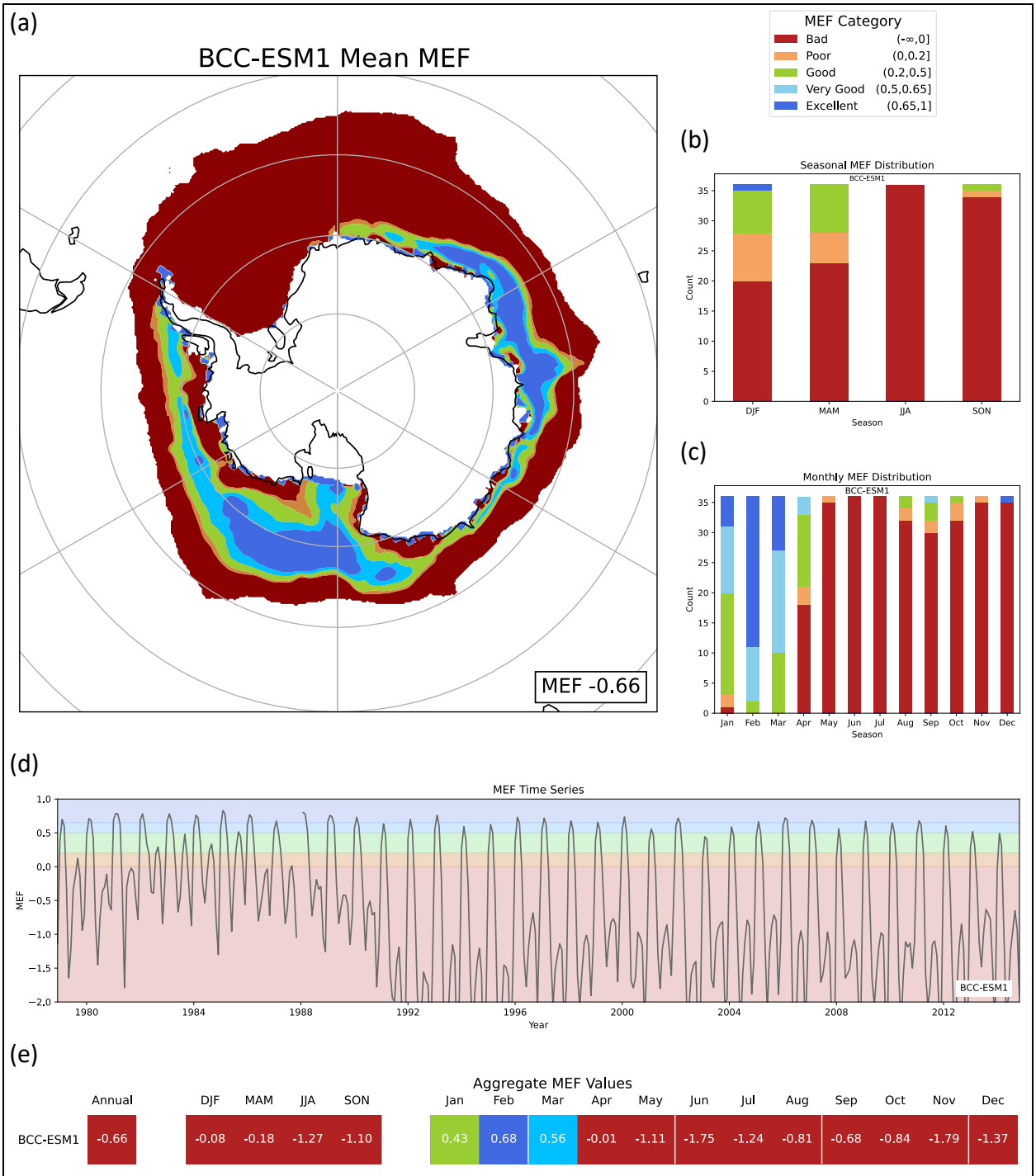


Figure B-23 BCC-ESM1 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

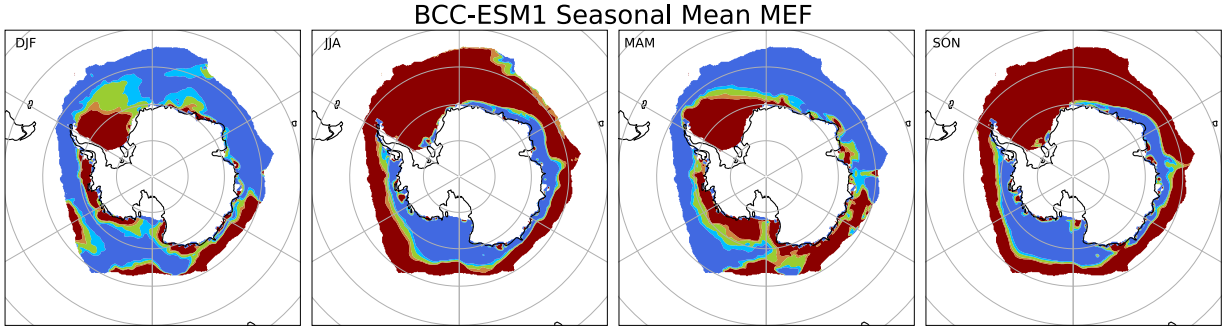


Figure B-24 BCC-ESM1 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

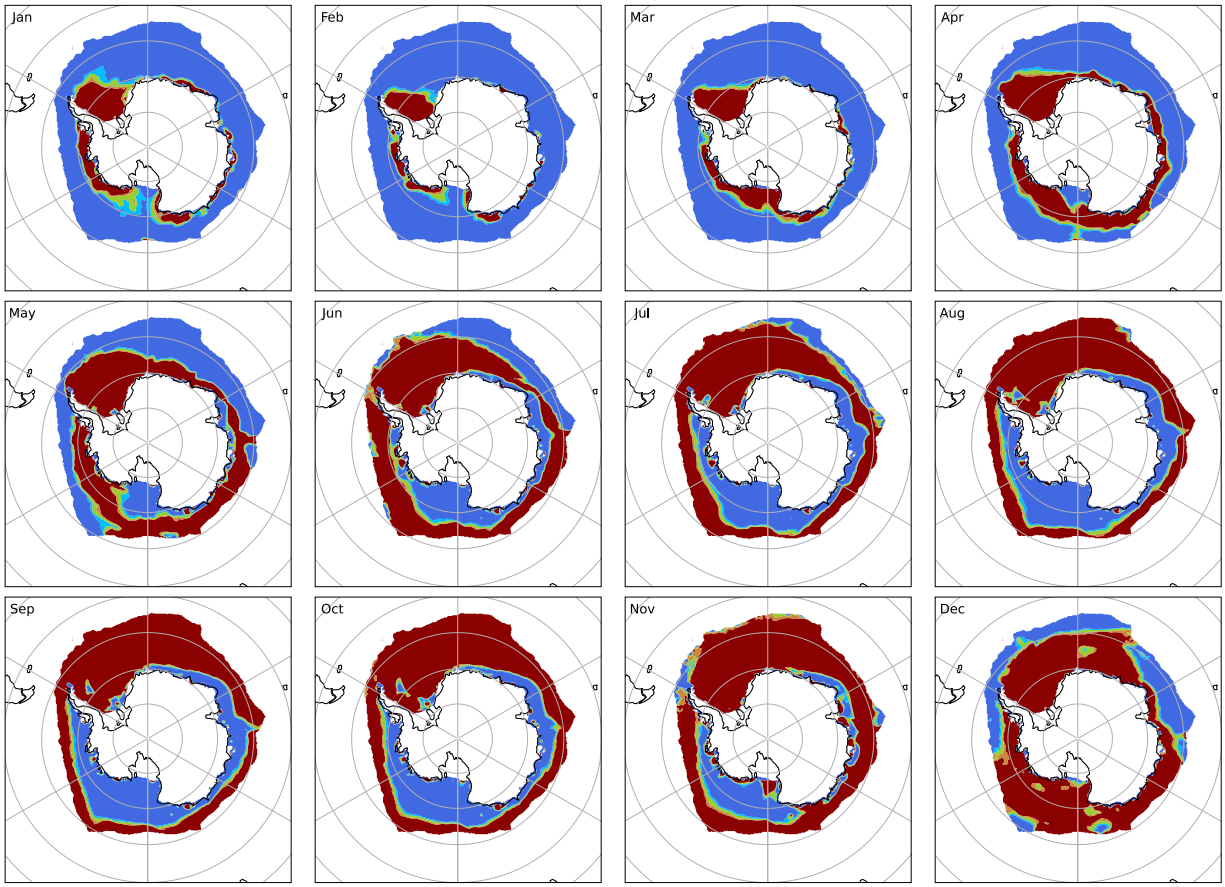


Figure B-25 BCC-ESM1 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

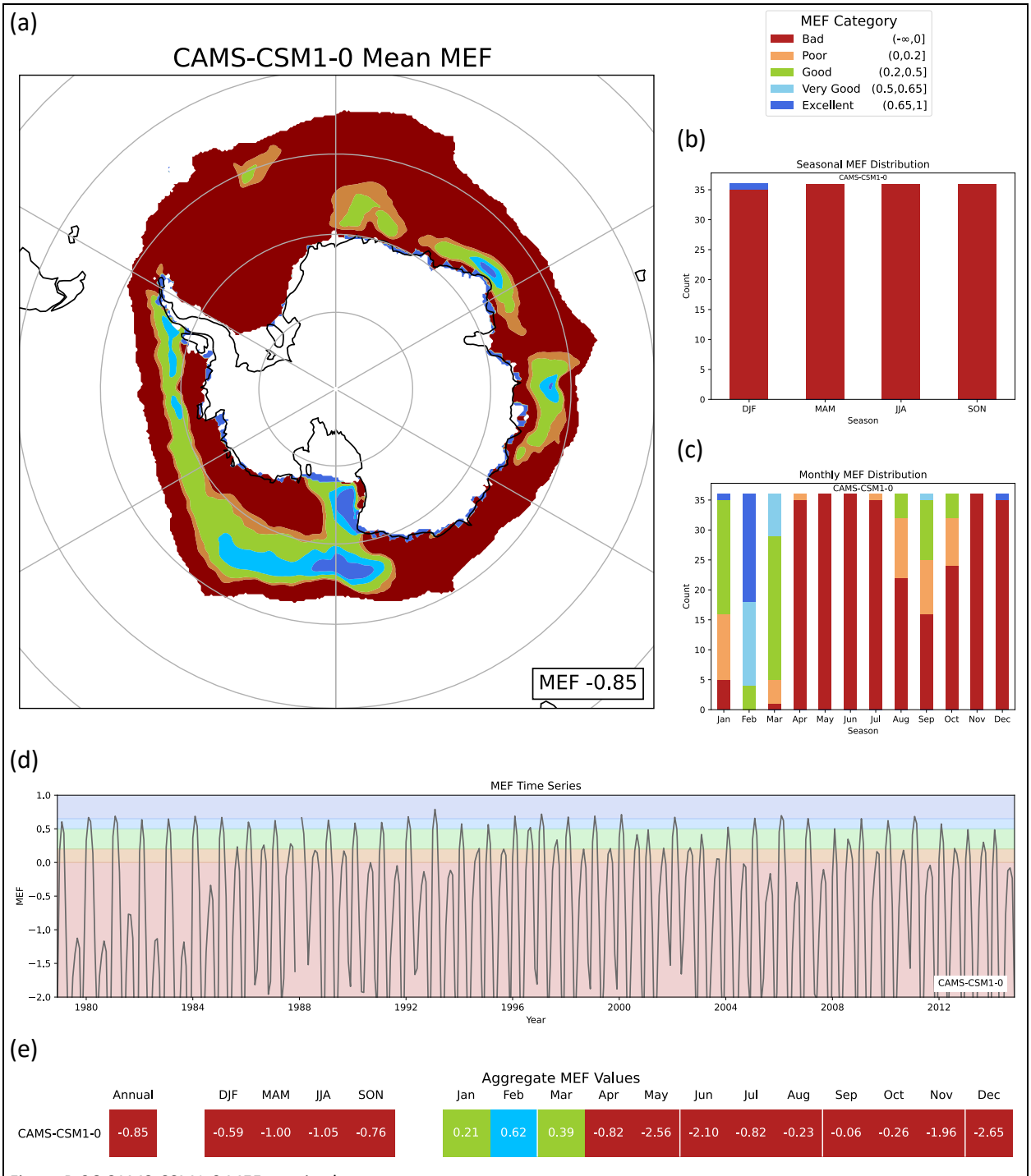


Figure B-26 CAMS-CSM1-0 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

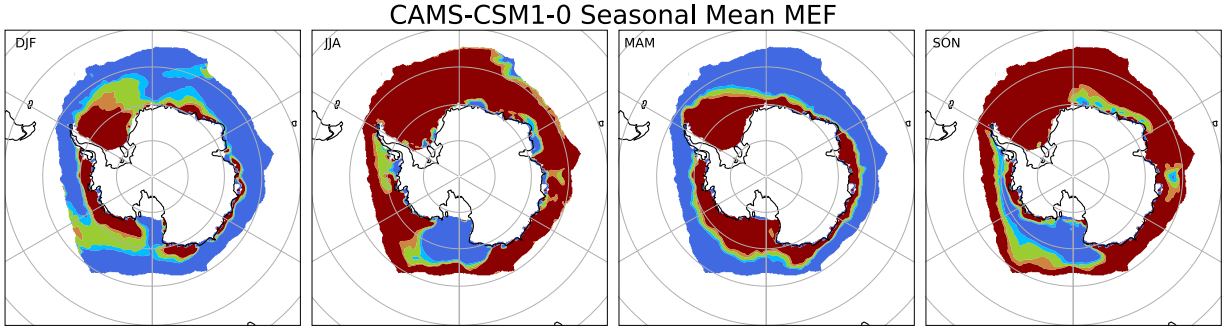


Figure B-27 CAMS-CSM1-0 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

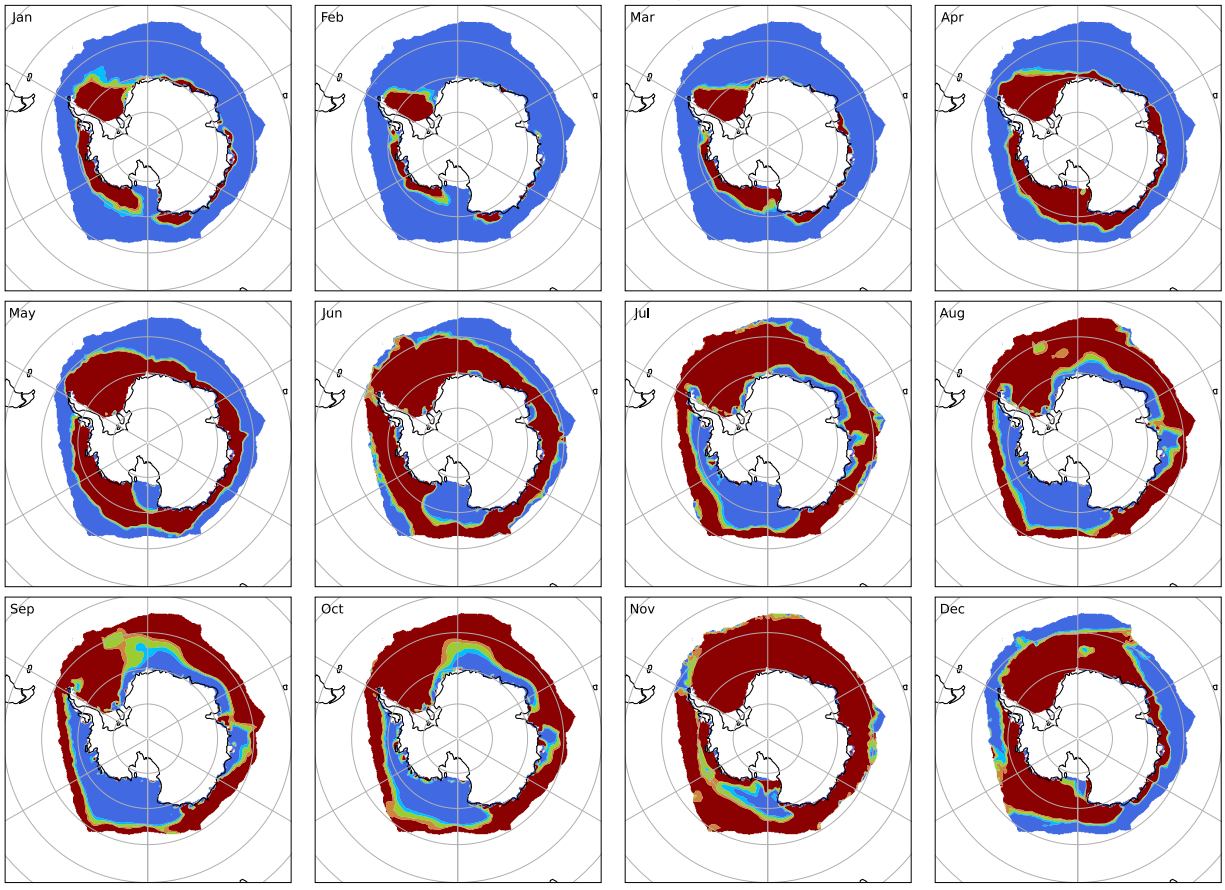


Figure B-28 CAMS-CSM1-0 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

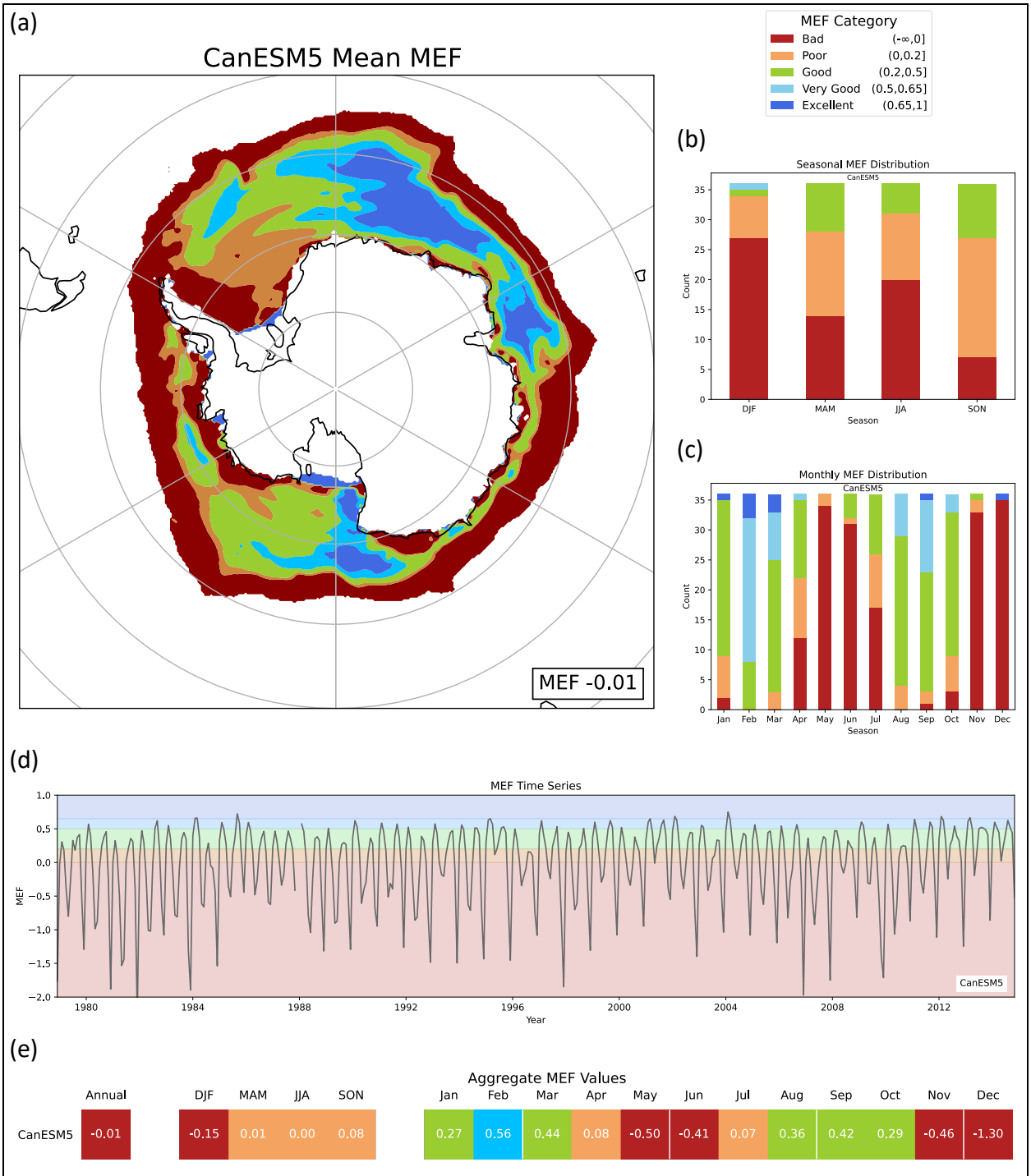


Figure B-29 CanESM5 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

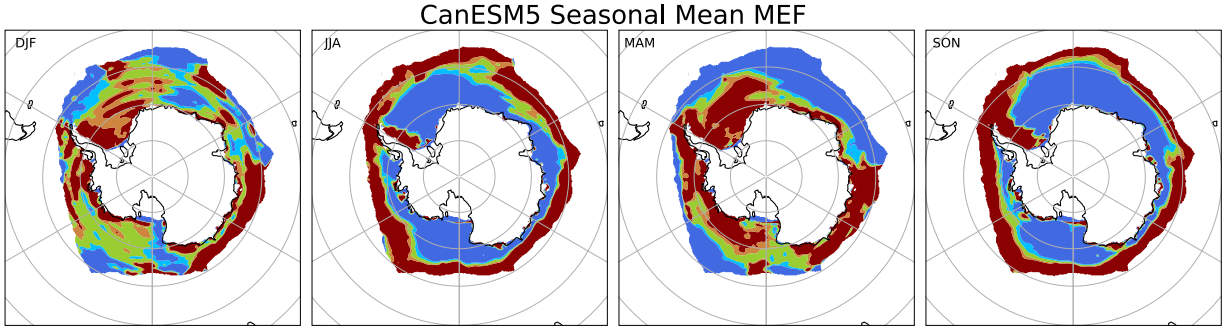


Figure B-30 CanESM5 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

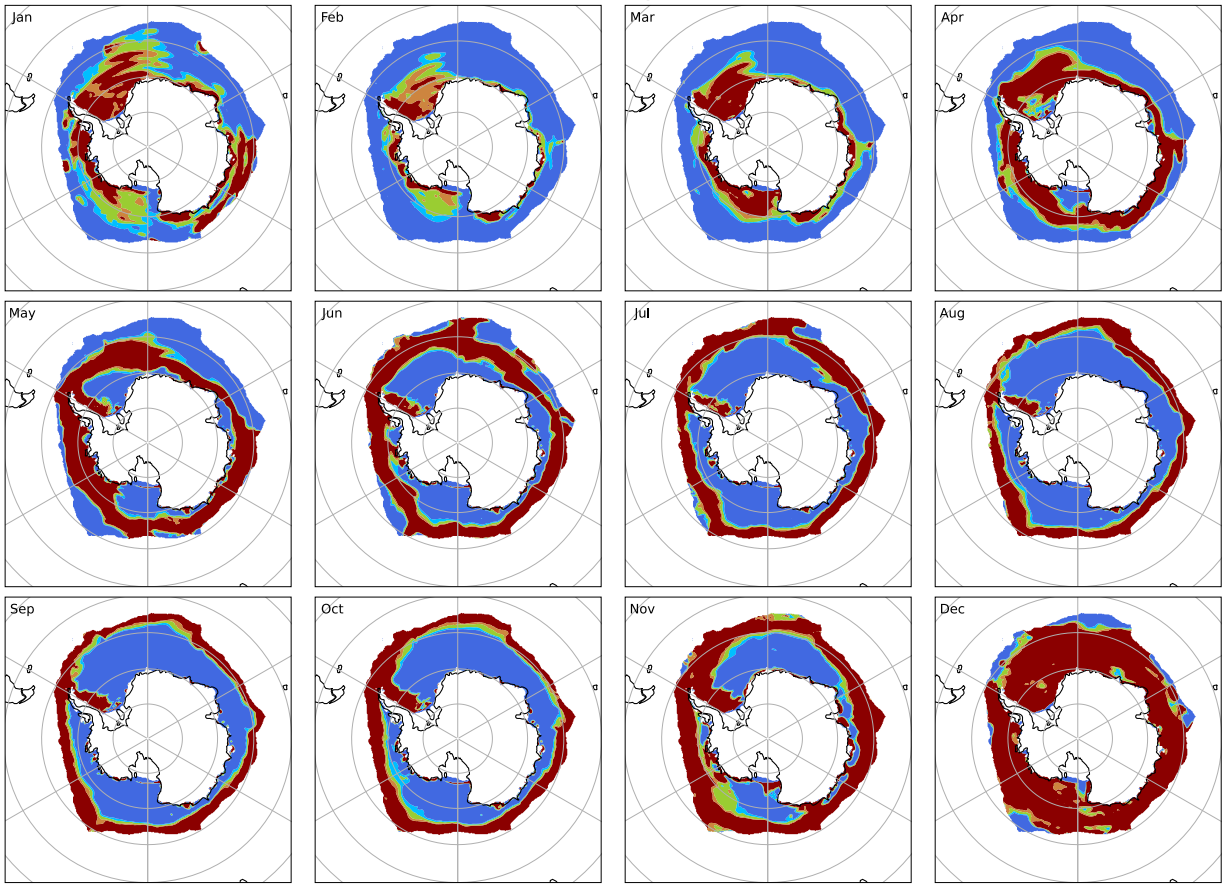


Figure B-31 CanESM5 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

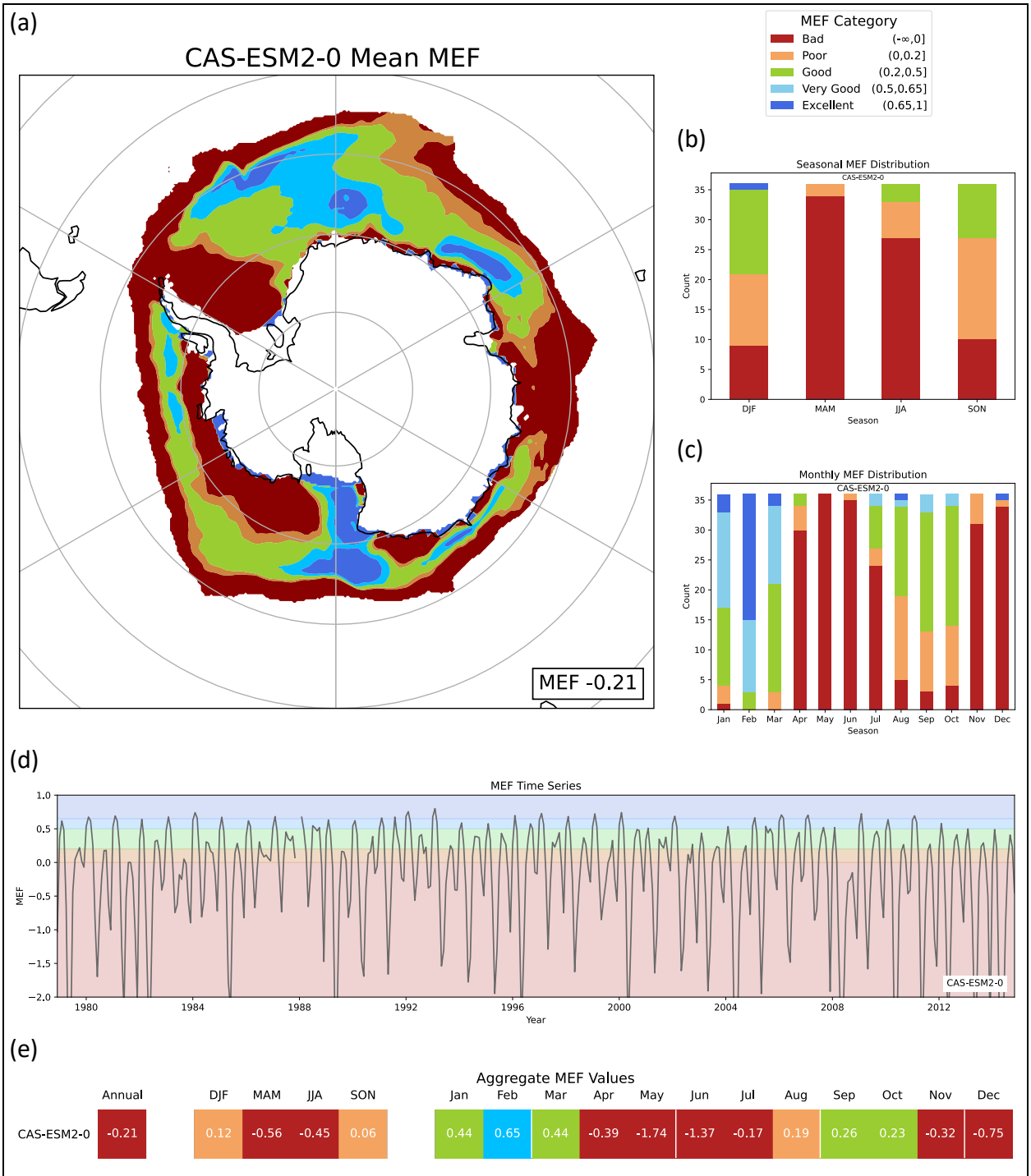


Figure B-32 CAS-ESM2-0 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

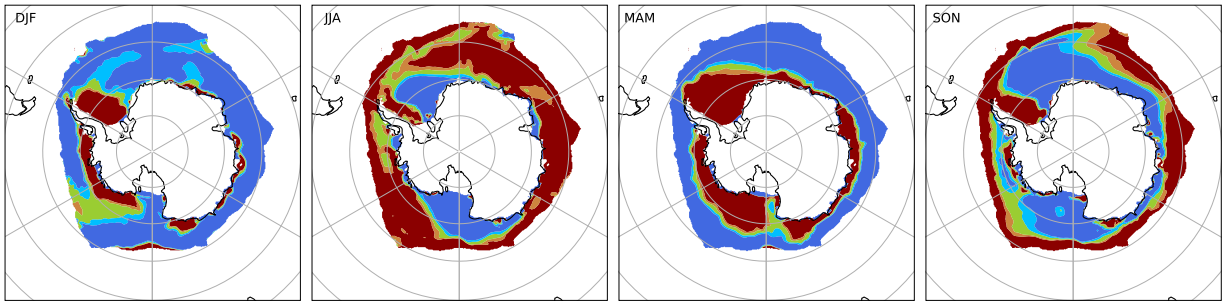


Figure B-33 CAS-ESM2-0 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

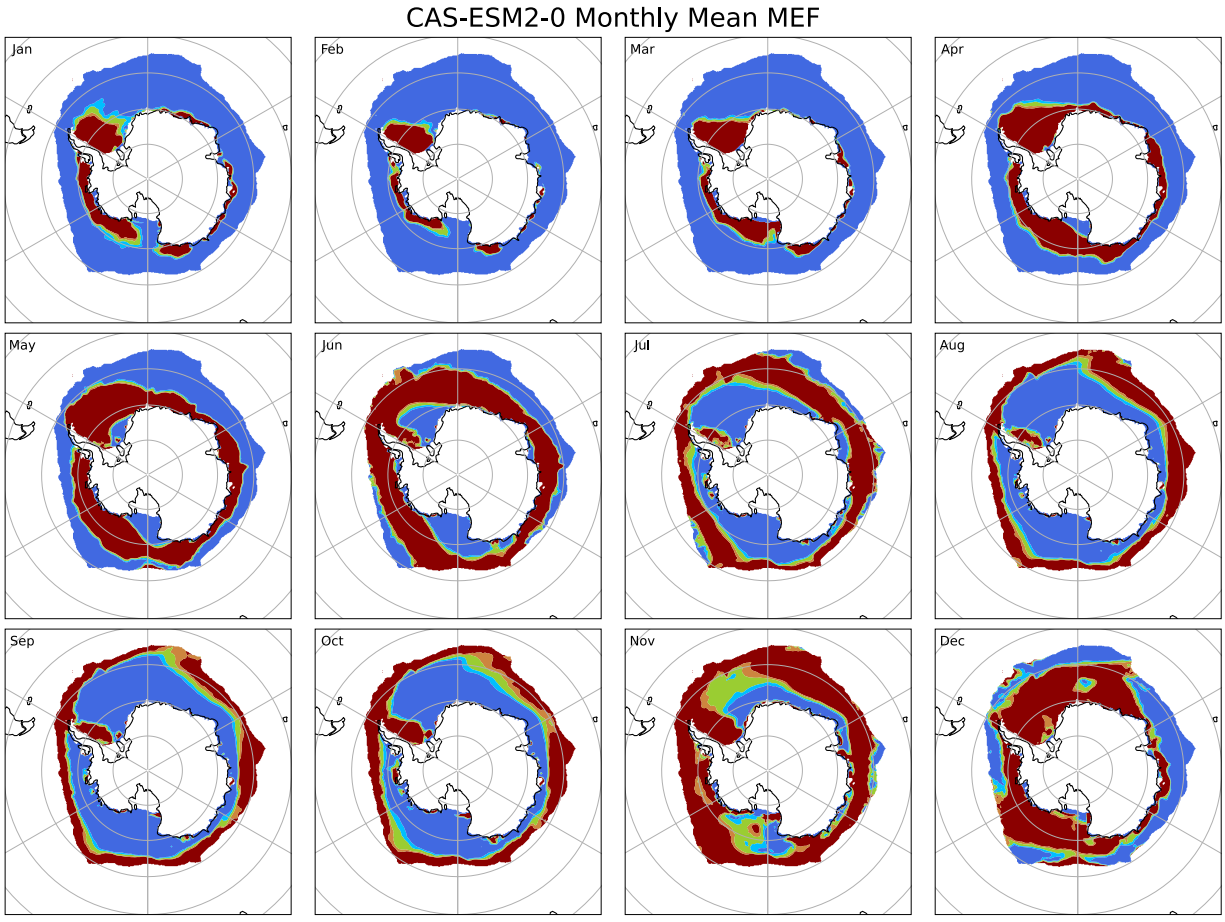


Figure B-34 CAS-ESM2-0 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

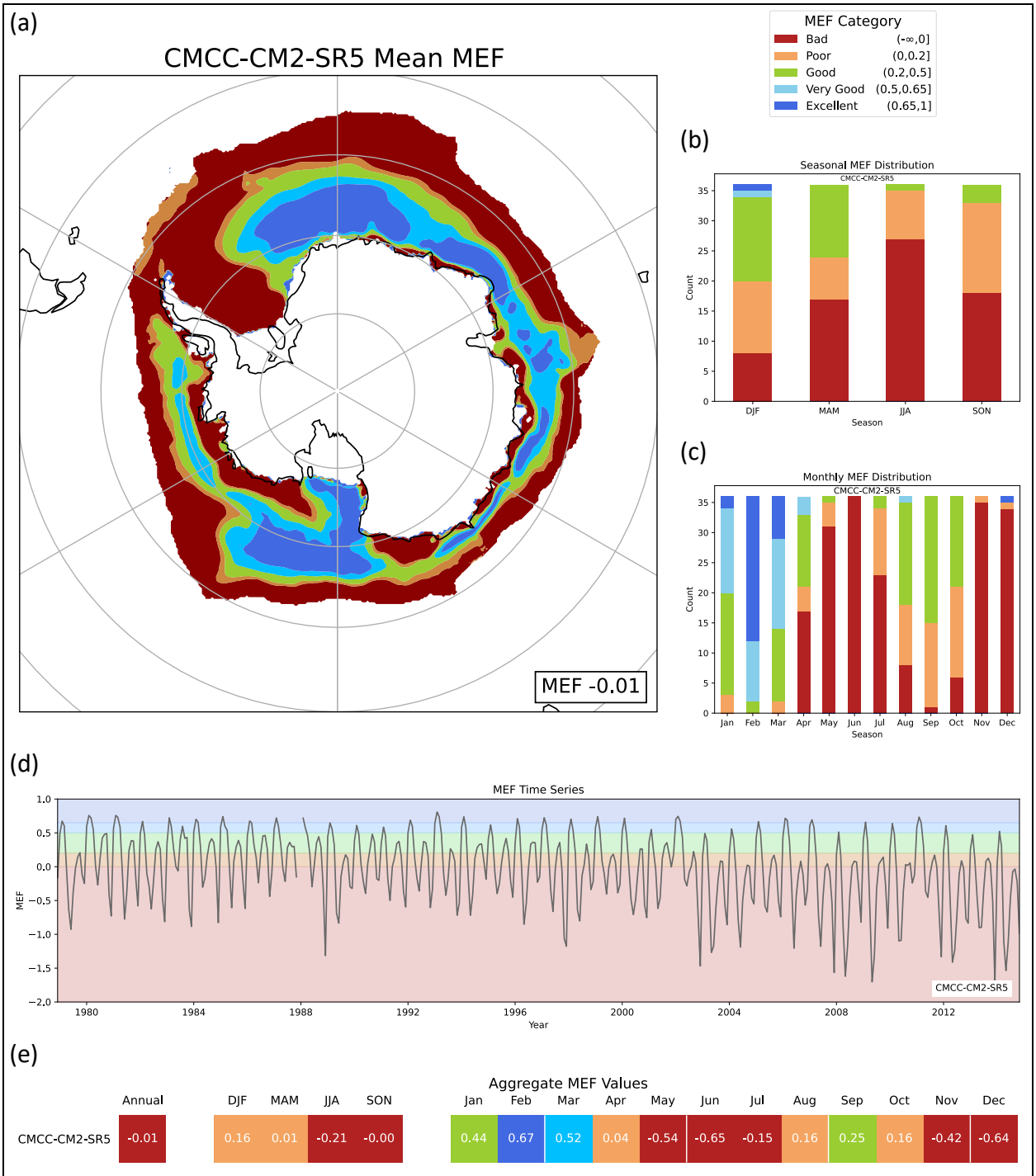


Figure B-35 CMCC-CM2-SR5 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

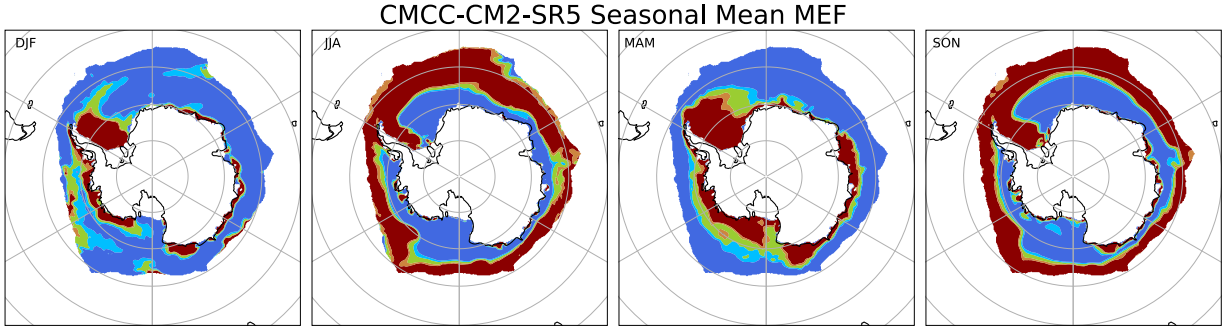


Figure B-36 CMCC-CM2-SR5 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

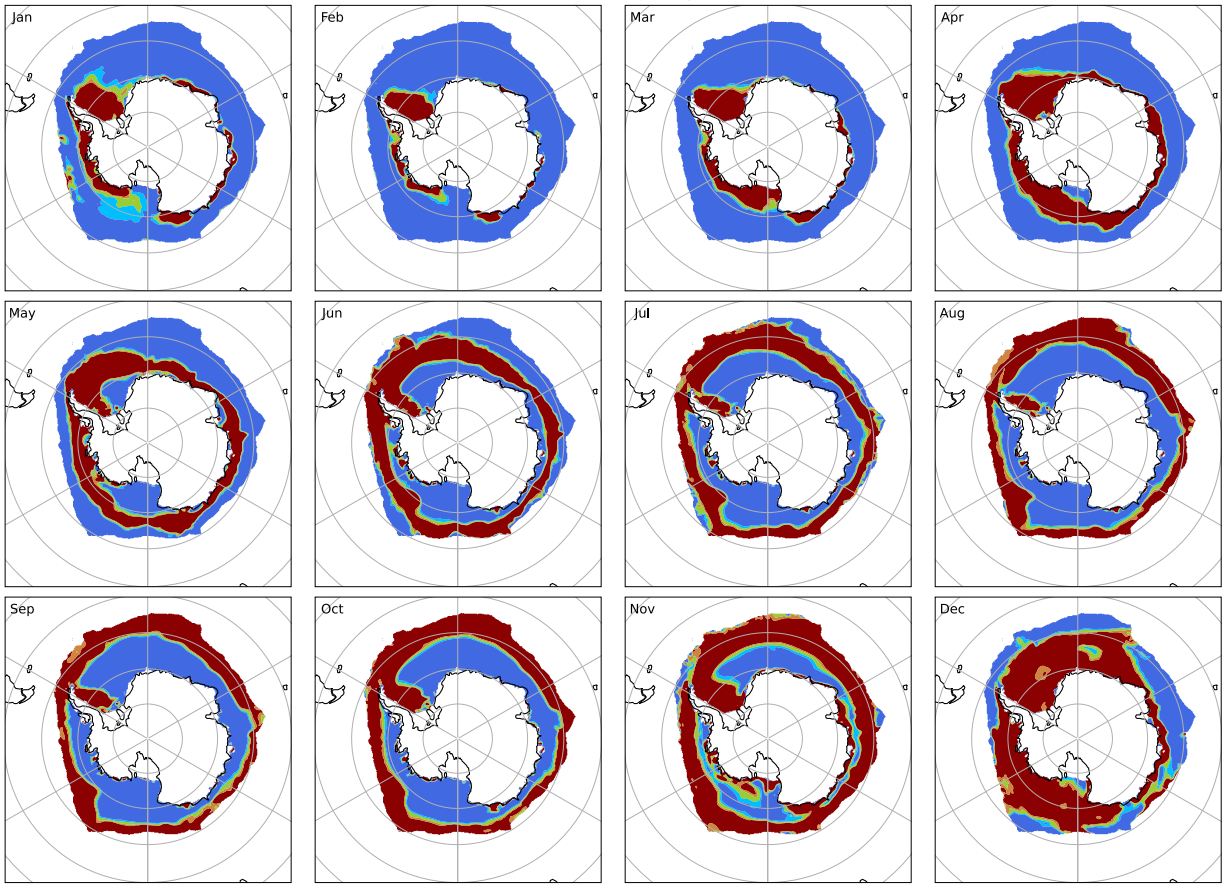


Figure B-37 CMCC-CM2-SR5 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

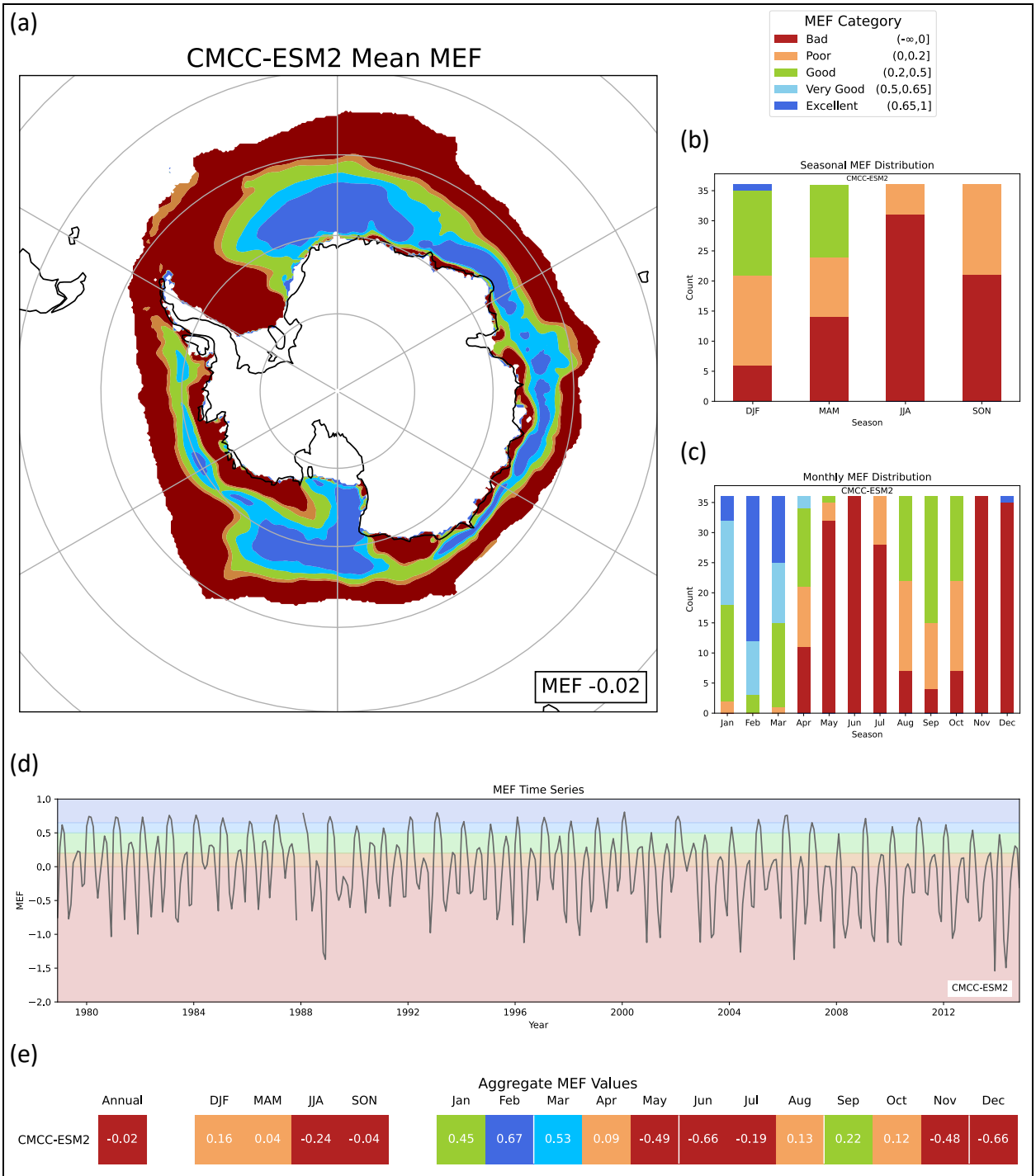


Figure B-38 CMCC-ESM2 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

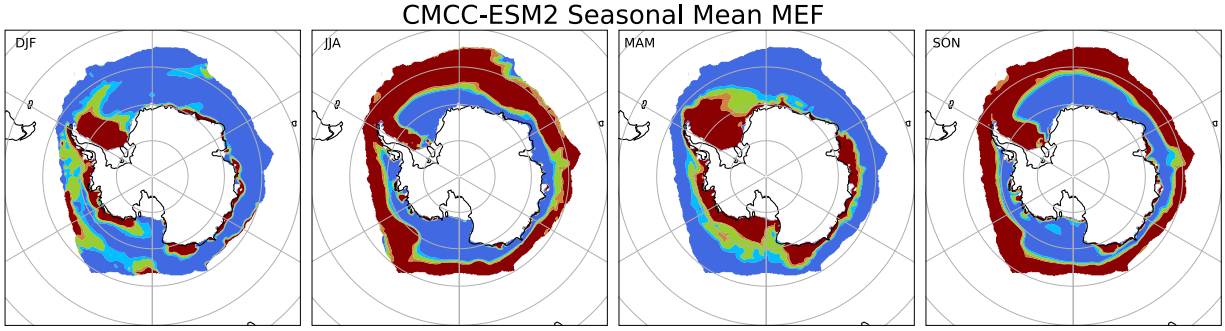


Figure B-39 CMCC-ESM2 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

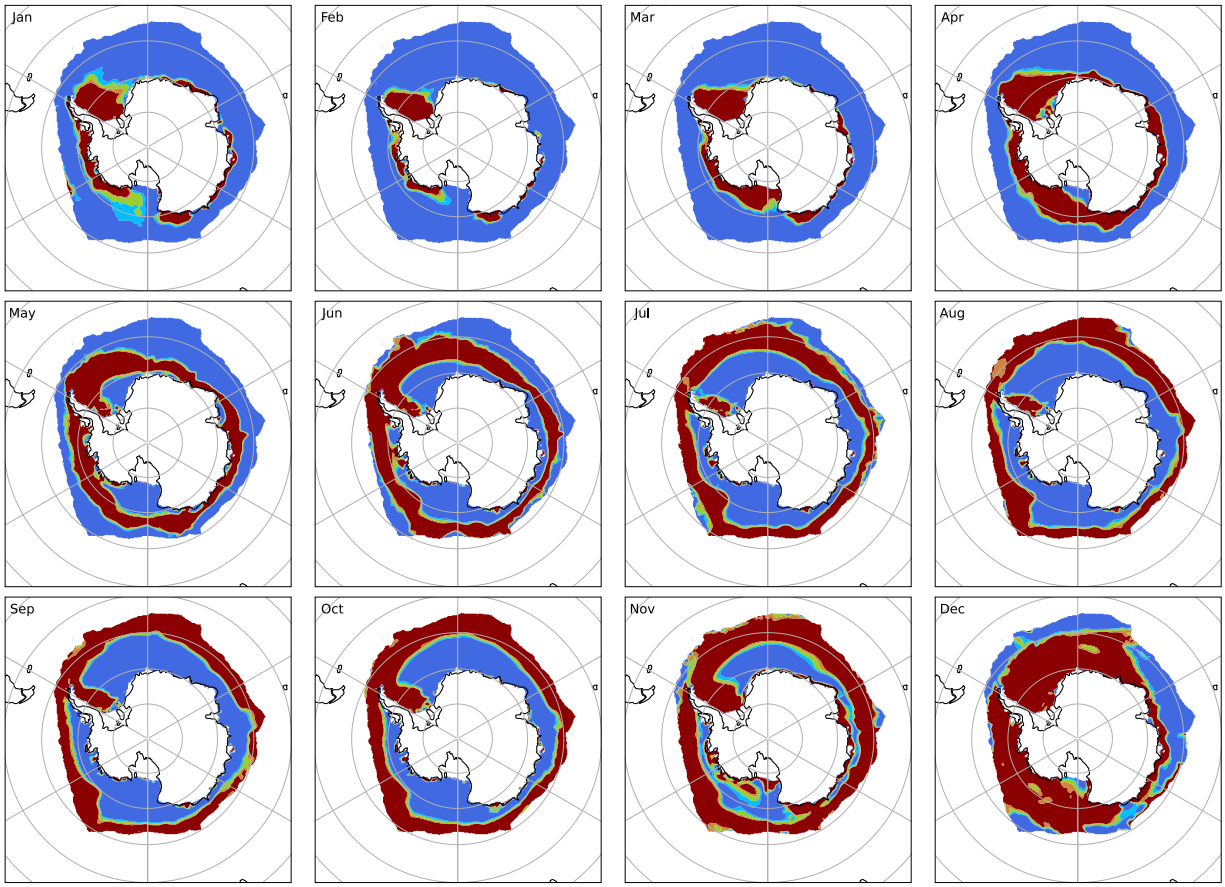


Figure B-40 CMCC-ESM2 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

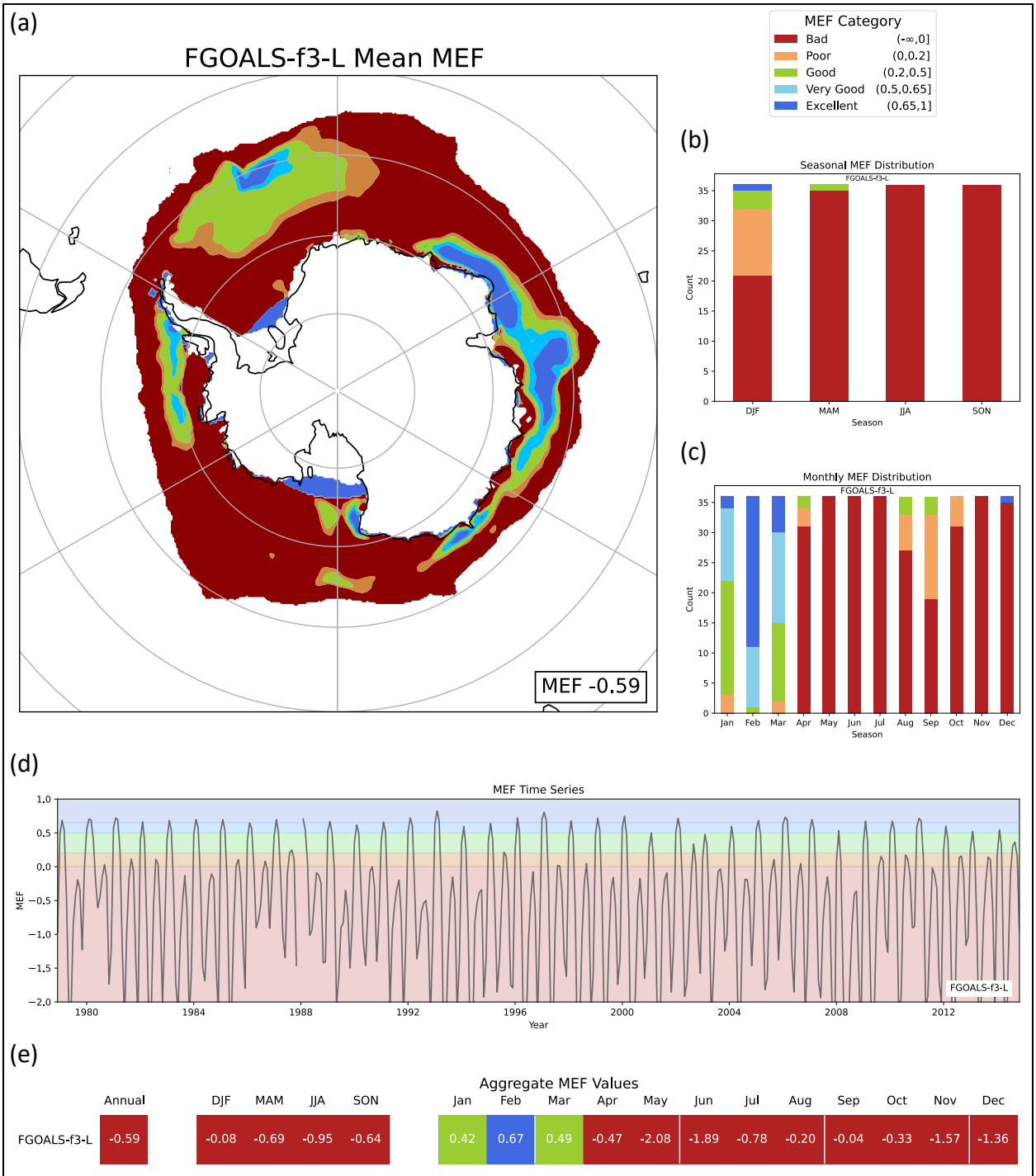


Figure B-41 FGOALS-f3-L MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

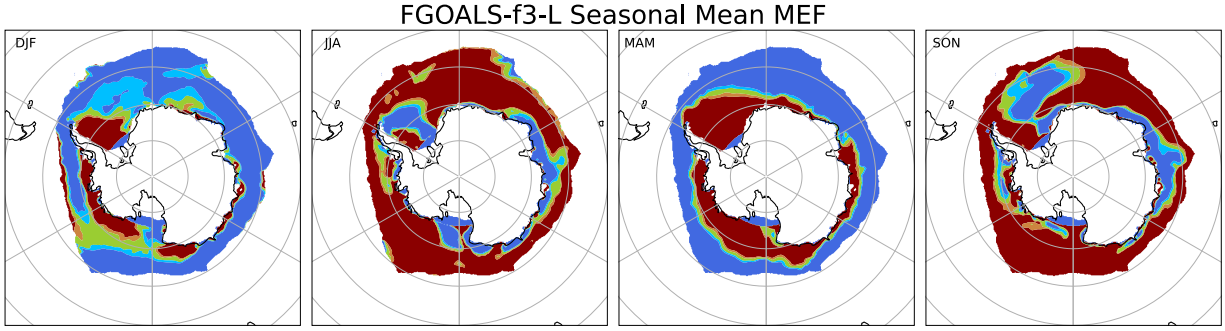


Figure B-42 FGOALS-f3-L seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

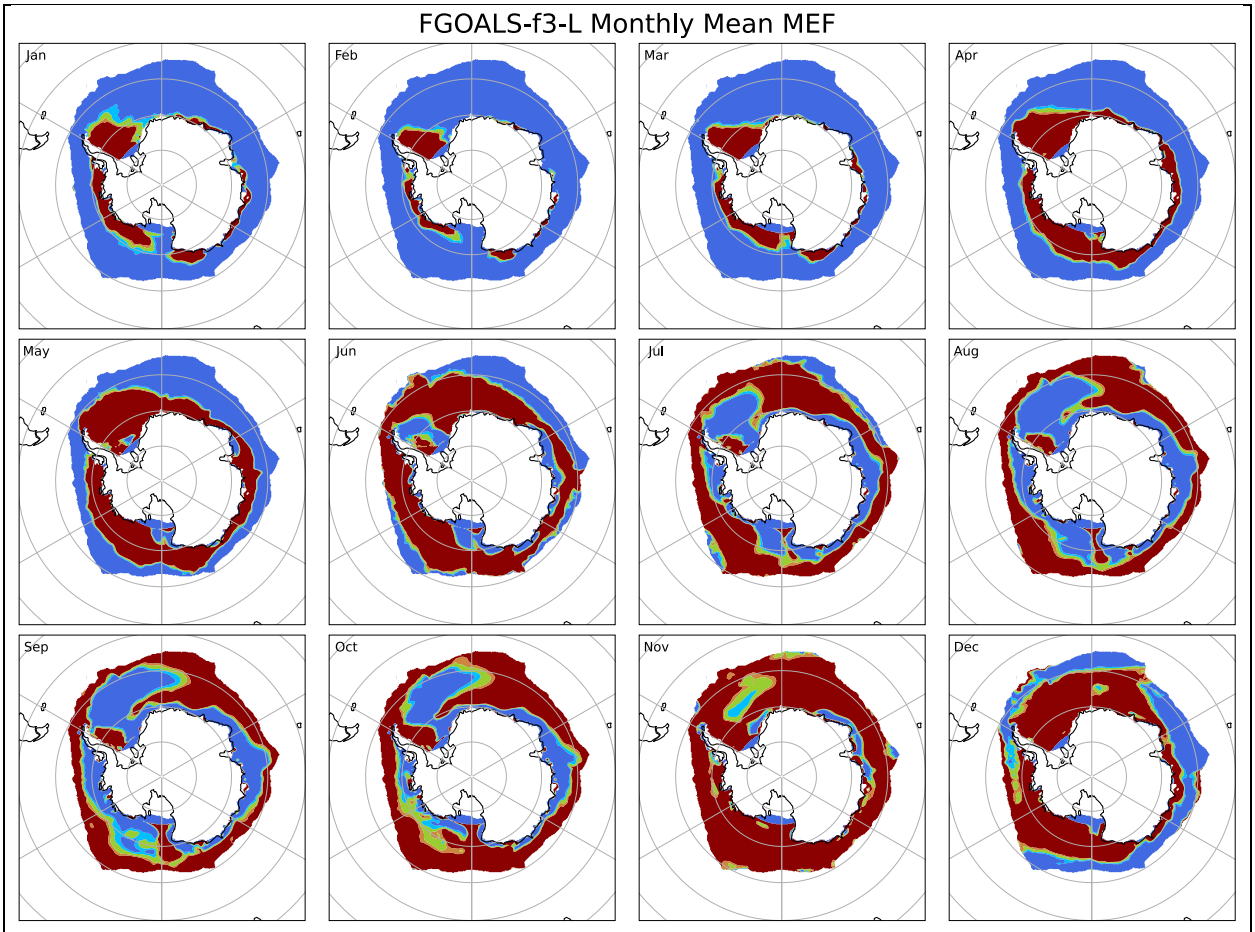


Figure B-43 FGOALS-f3-L monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

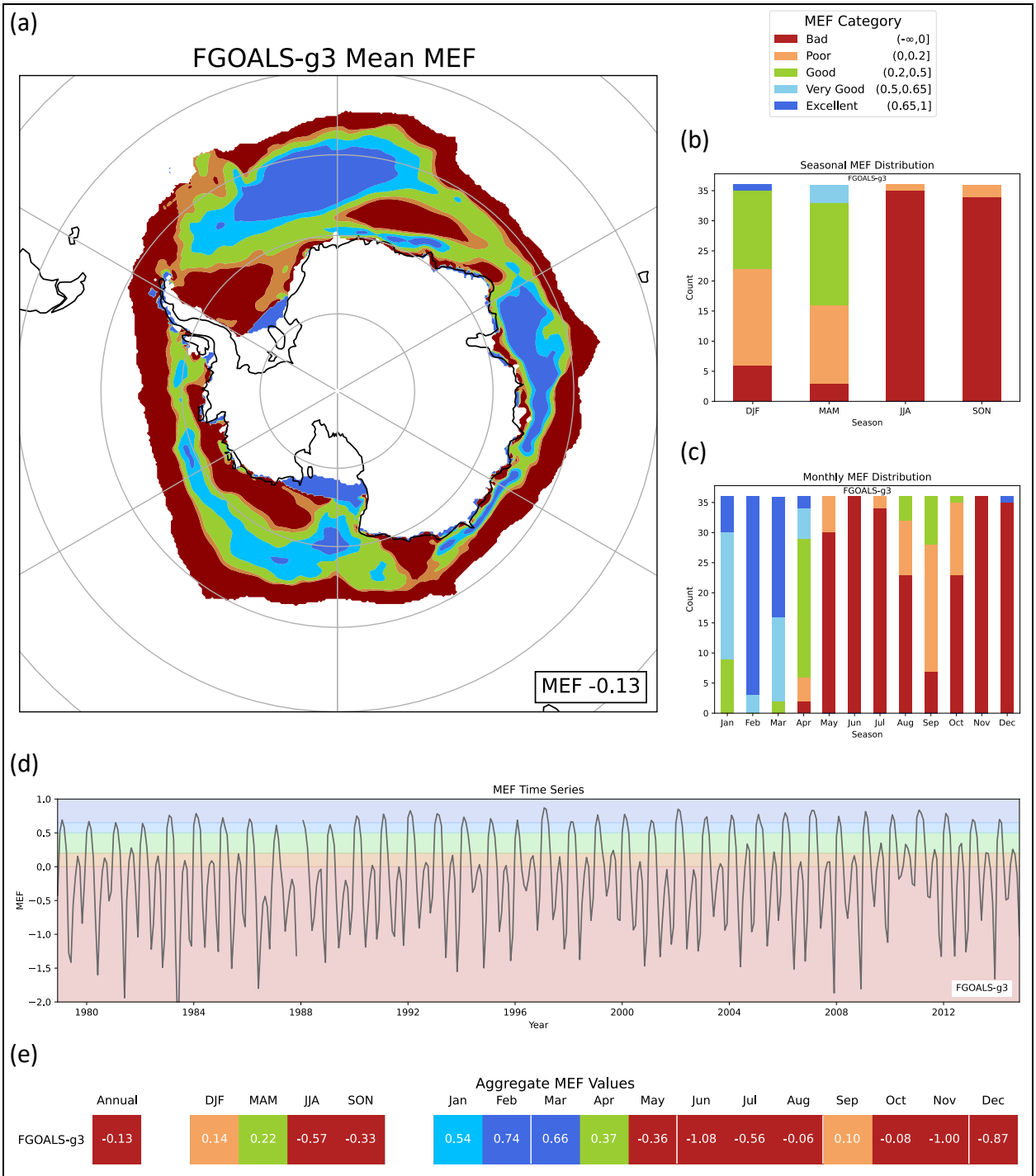


Figure B-44 FGOALS-g3 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

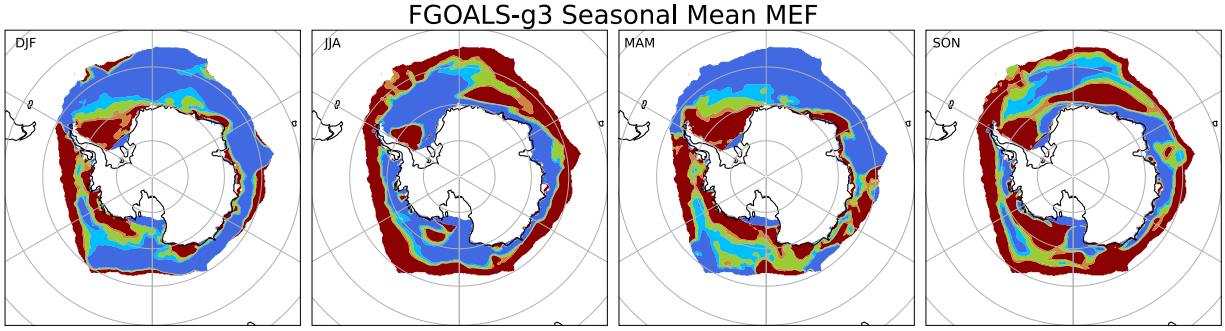


Figure B-45 FGOALS-g3 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

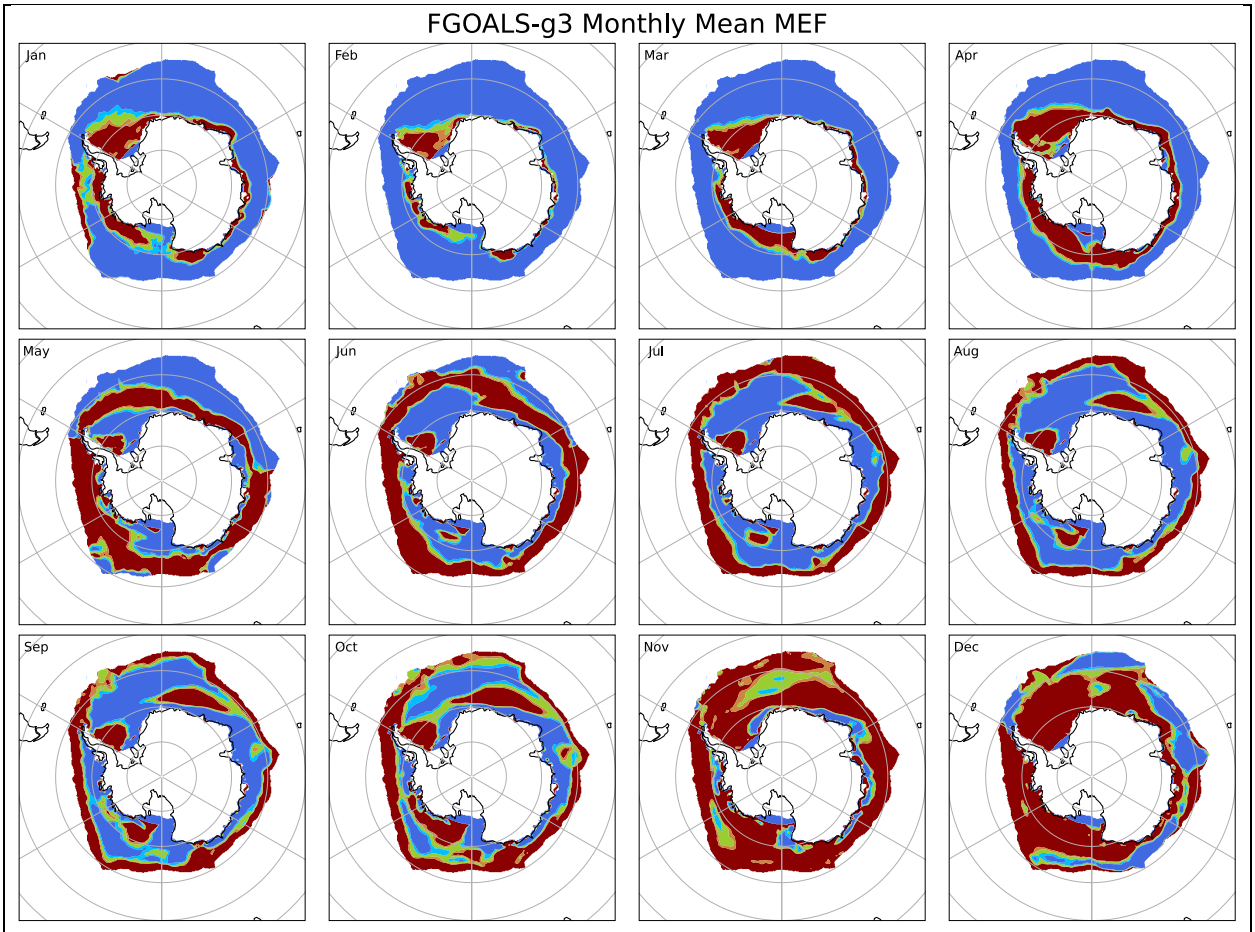


Figure B-46 FGOALS-g3 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

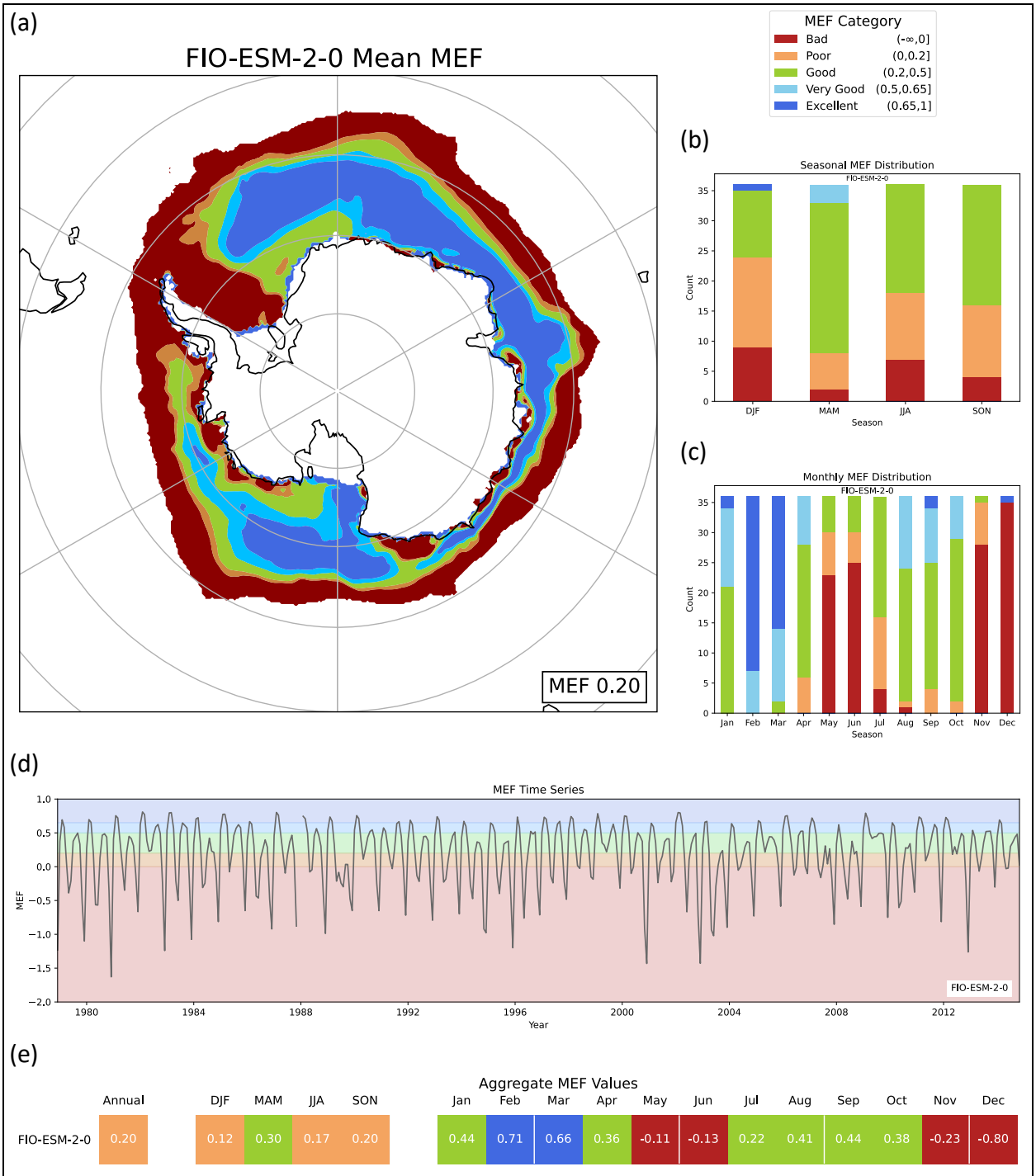


Figure B-47 FIO-ESM-2-0 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

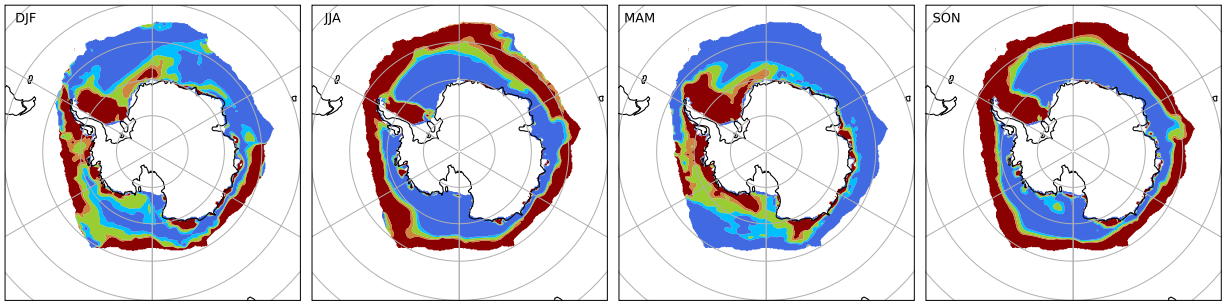


Figure B-48 FIO-ESM-2-0 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

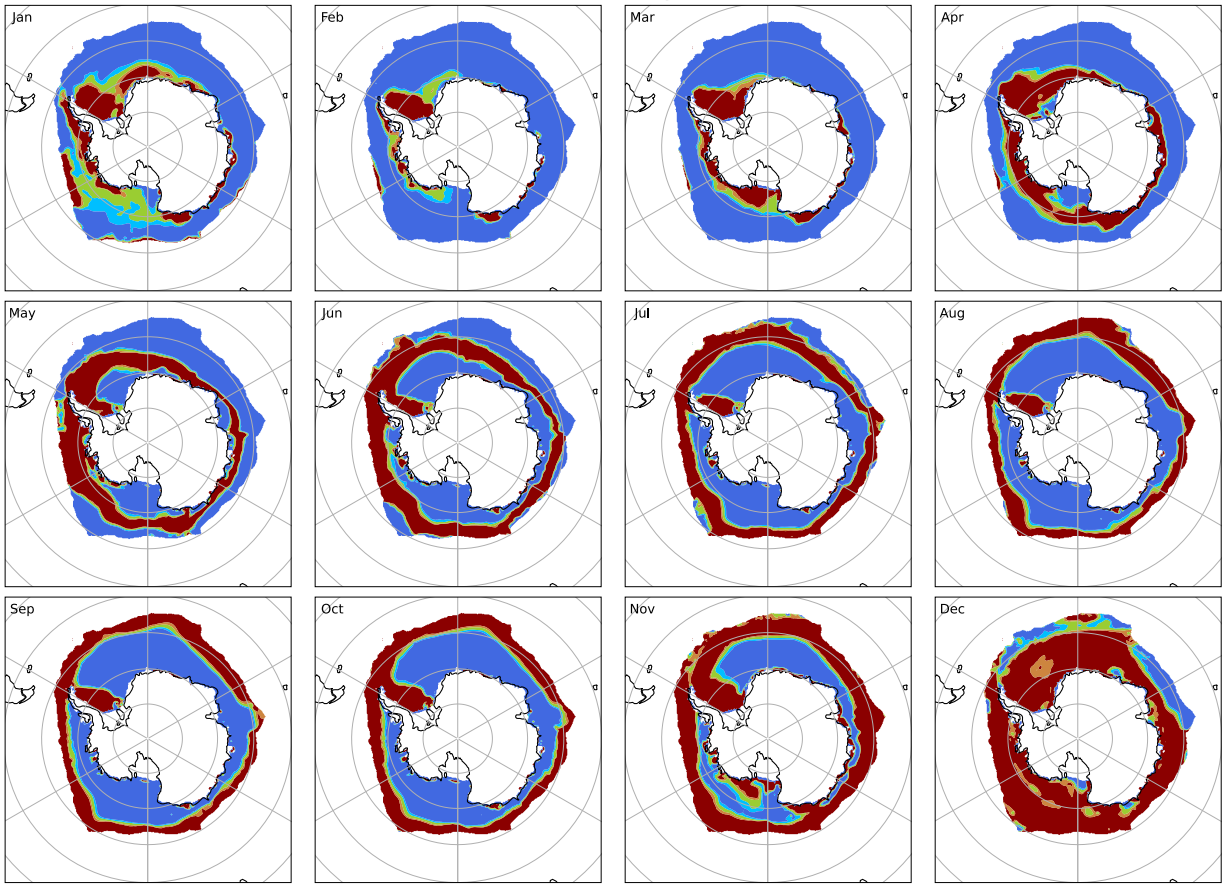


Figure B-49 FIO-ESM-2-0 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

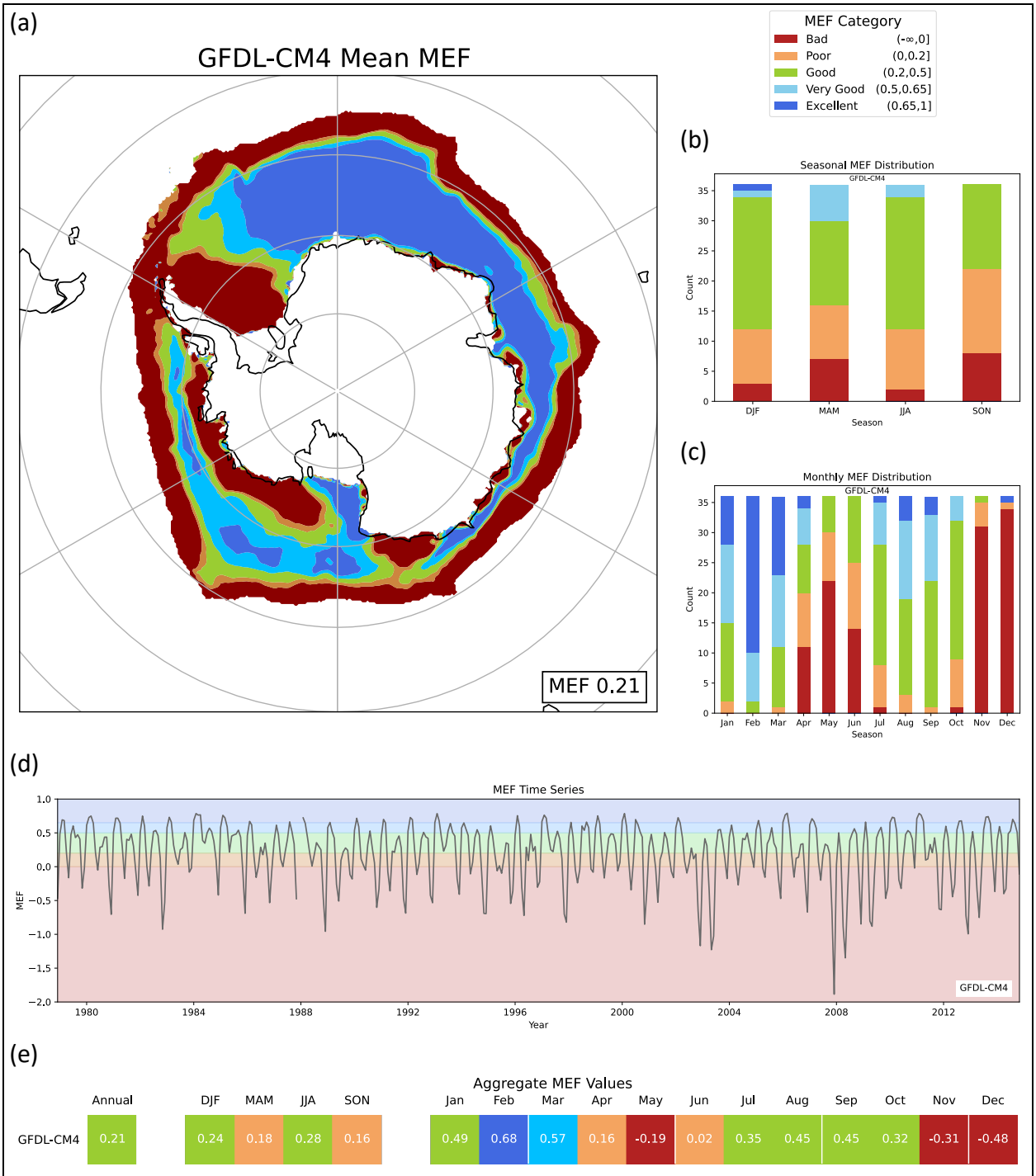


Figure B-50 GFDL-CM4 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

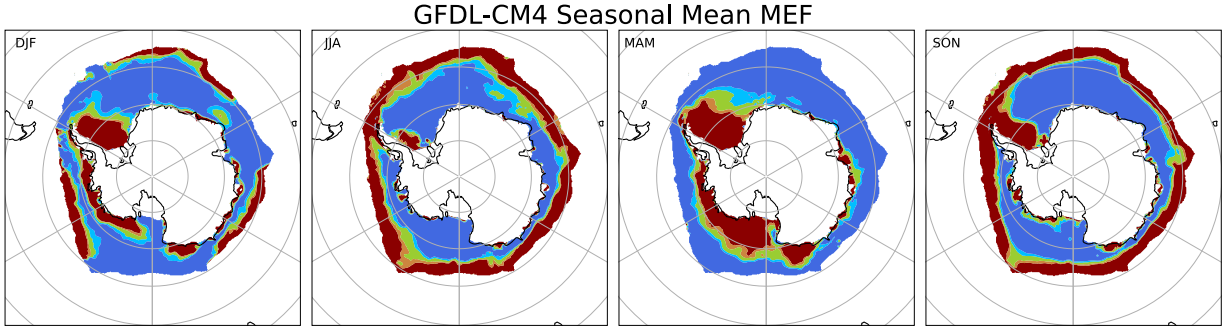


Figure B-51 GFDL-CM4 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

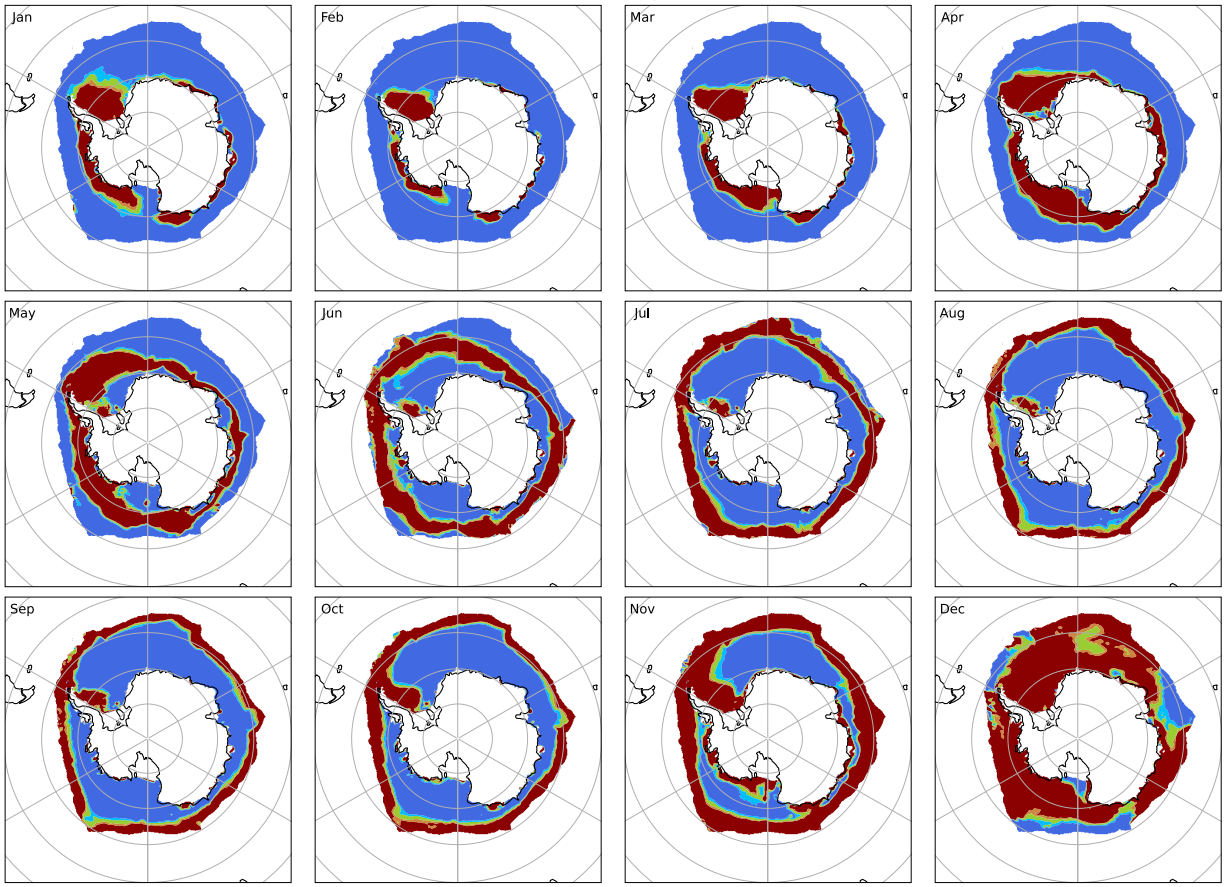


Figure B-52 GFDL-CM4 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

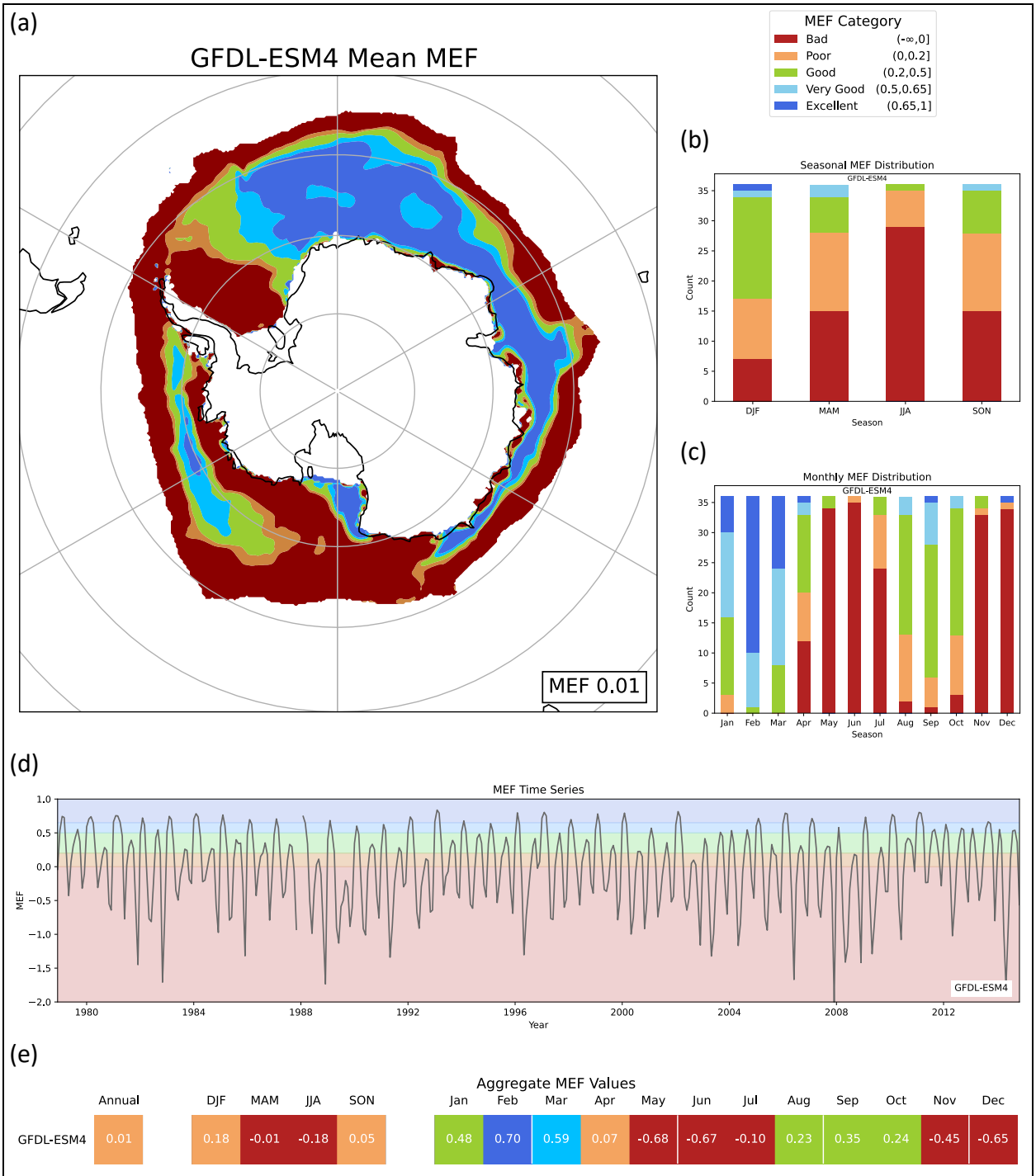


Figure B-53 GFDL-ESM4 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

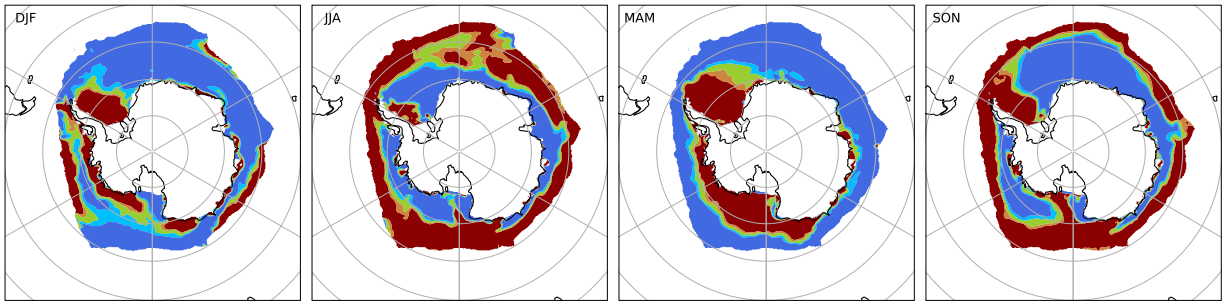


Figure B-54 GFDL-ESM4 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

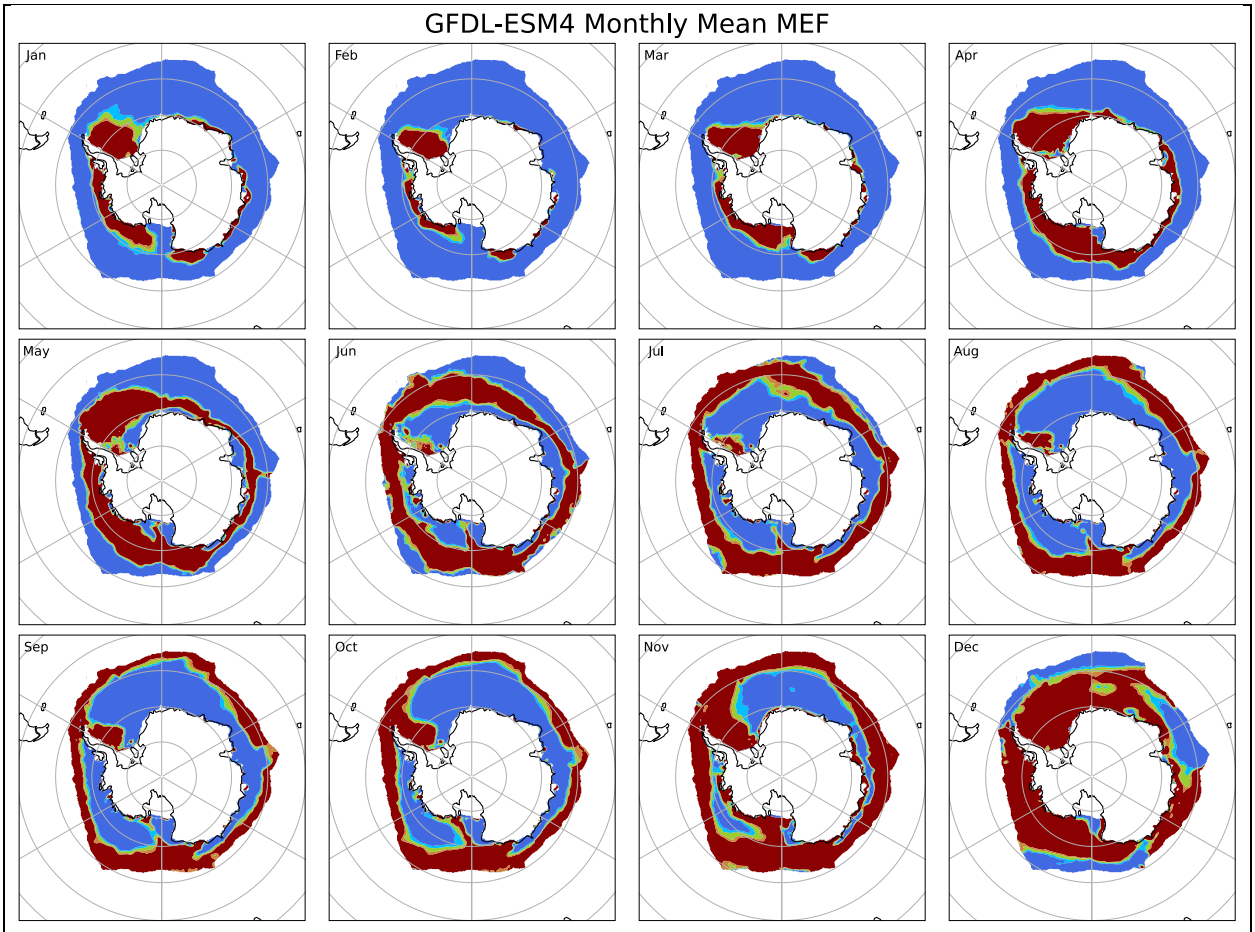


Figure B-55 GFDL-ESM4 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

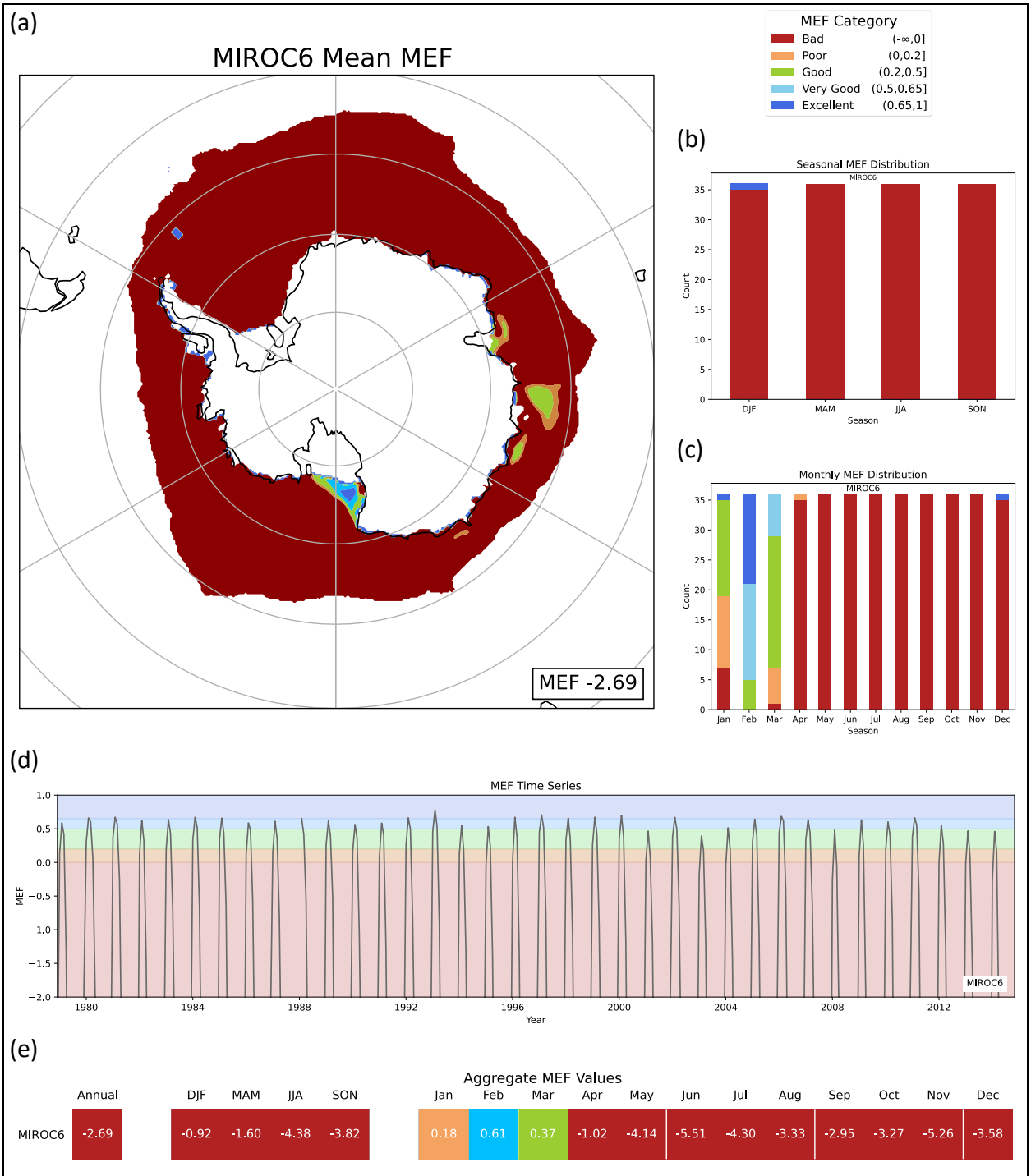


Figure B-56 MIROC6 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

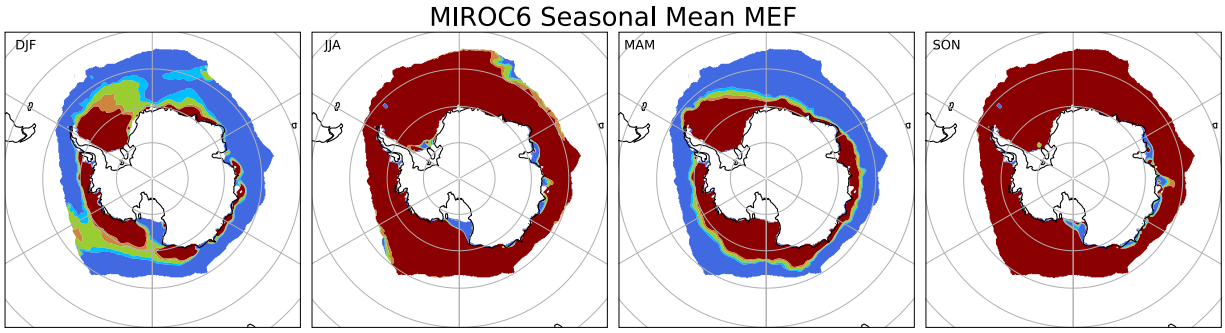


Figure B-57 MIROC6 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

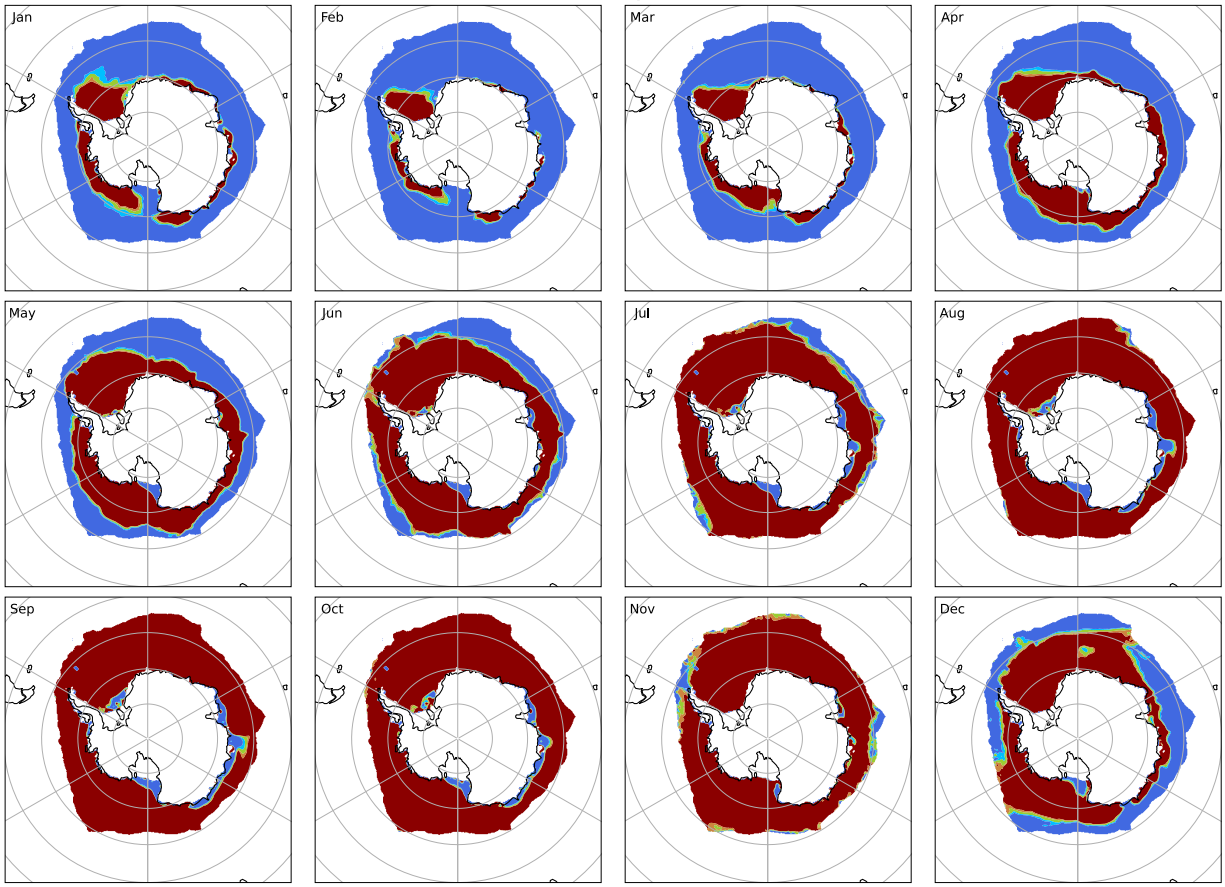


Figure B-58 MIROC6 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

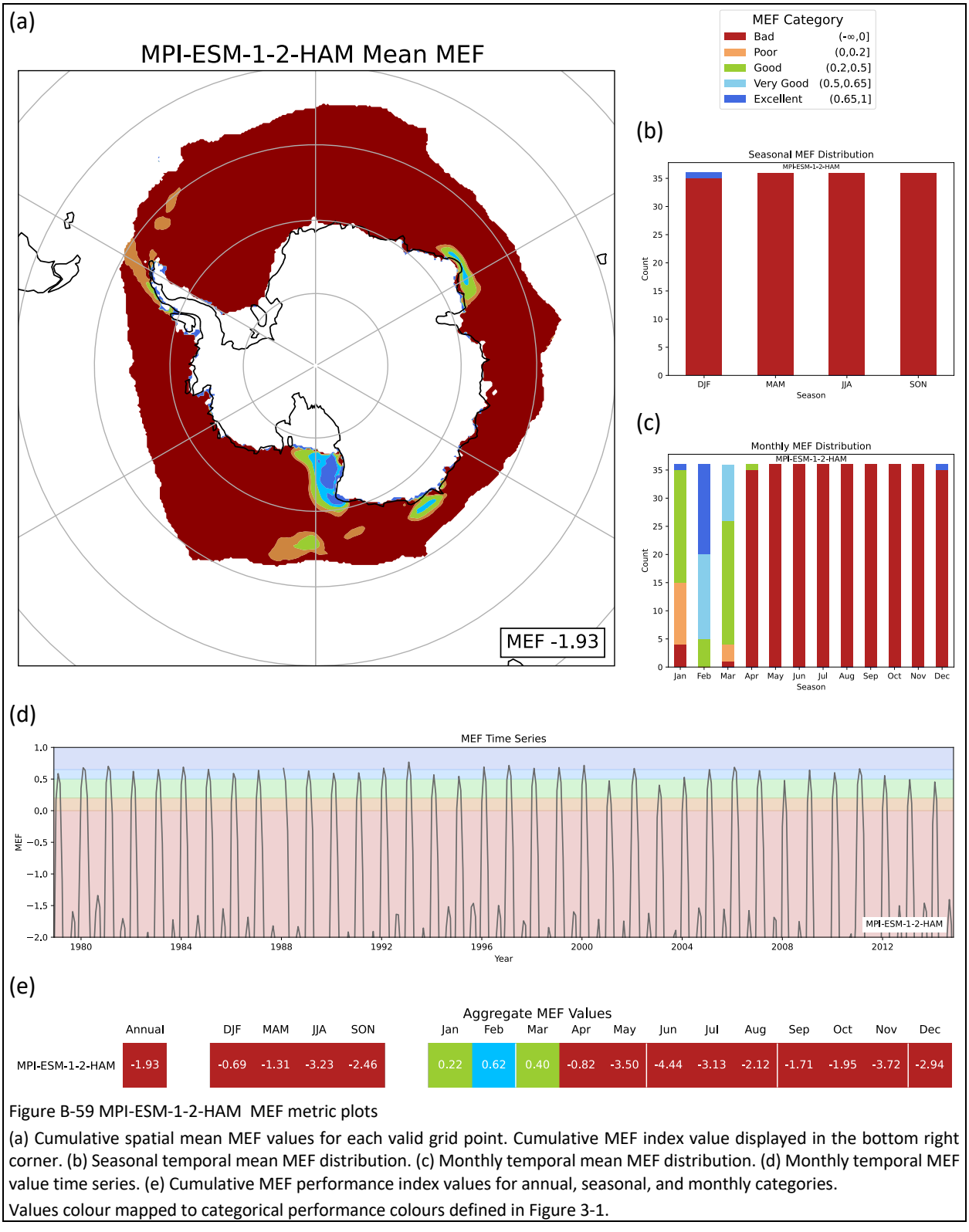


Figure B-59 MPI-ESM-1-2-HAM MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

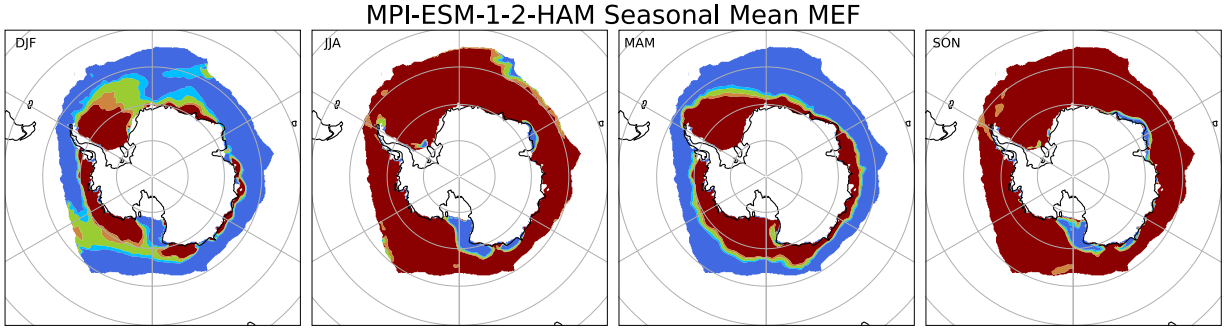


Figure B-60 MPI-ESM-1-2-HAM seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

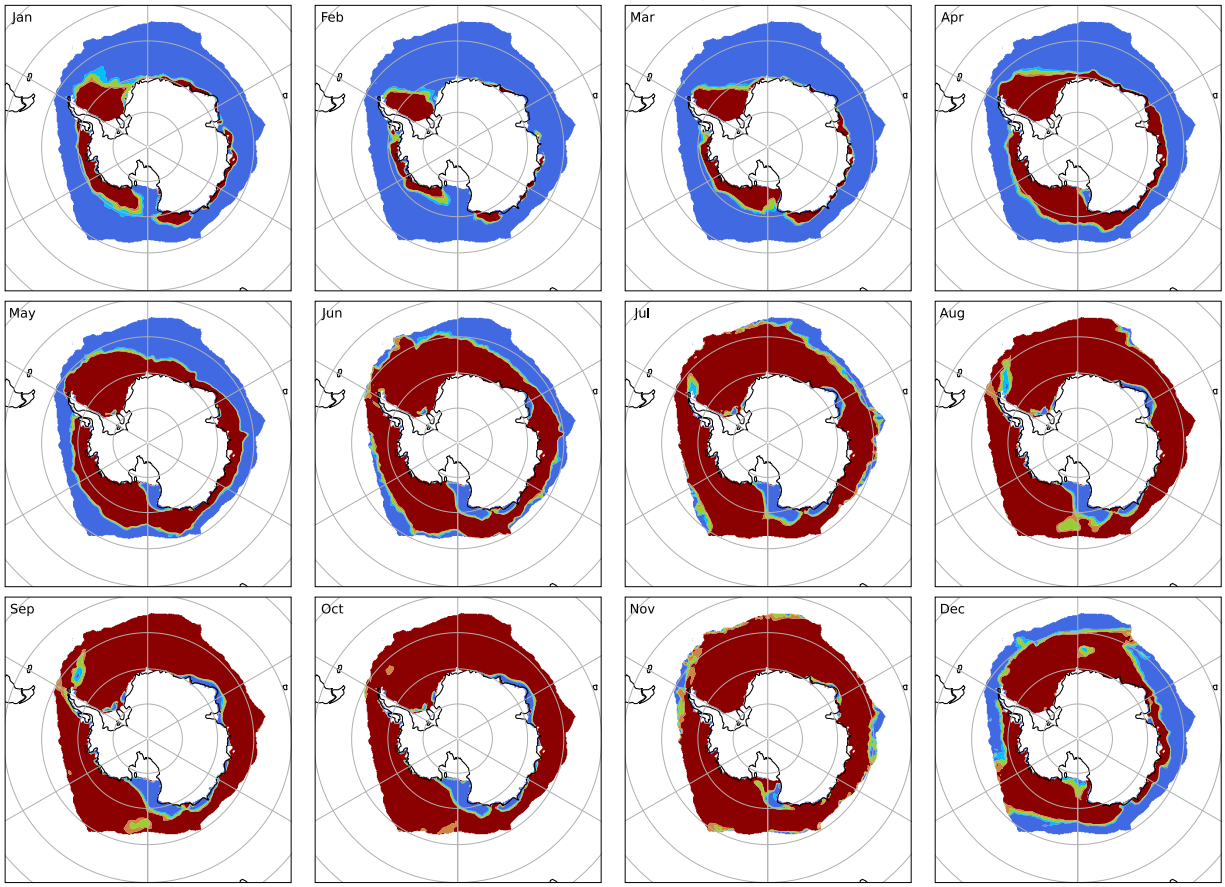


Figure B-61 MPI-ESM-1-2-HAM monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

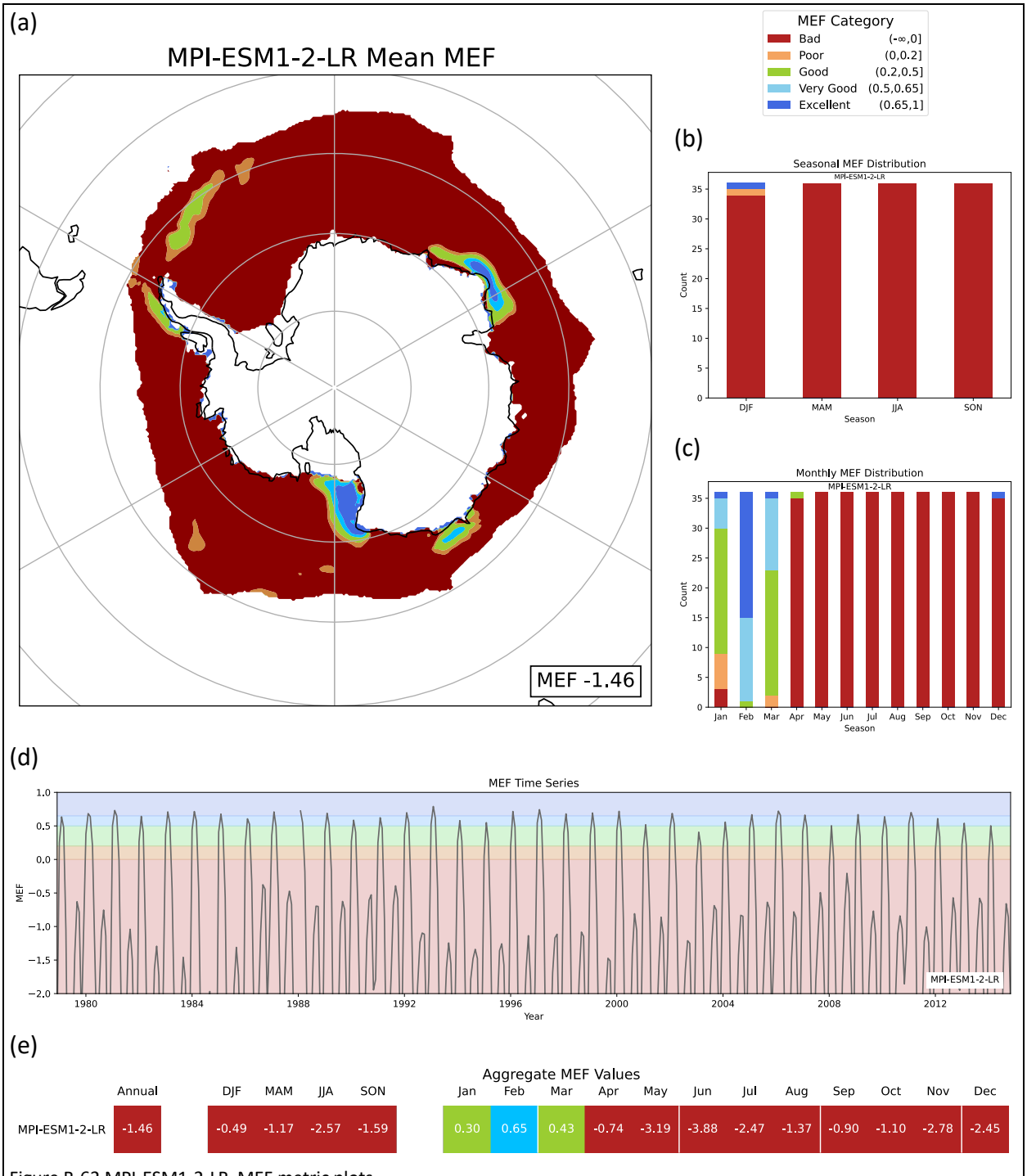


Figure B-62 MPI-ESM1-2-LR MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

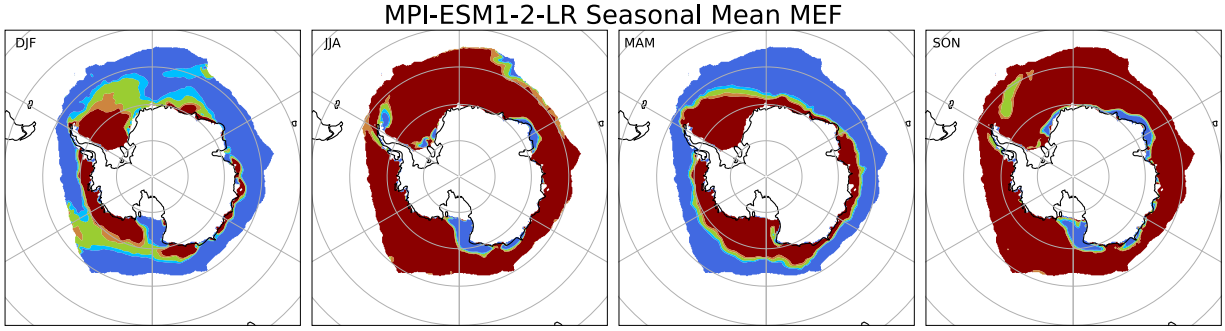


Figure B-63 MPI-ESM1-2-LR seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

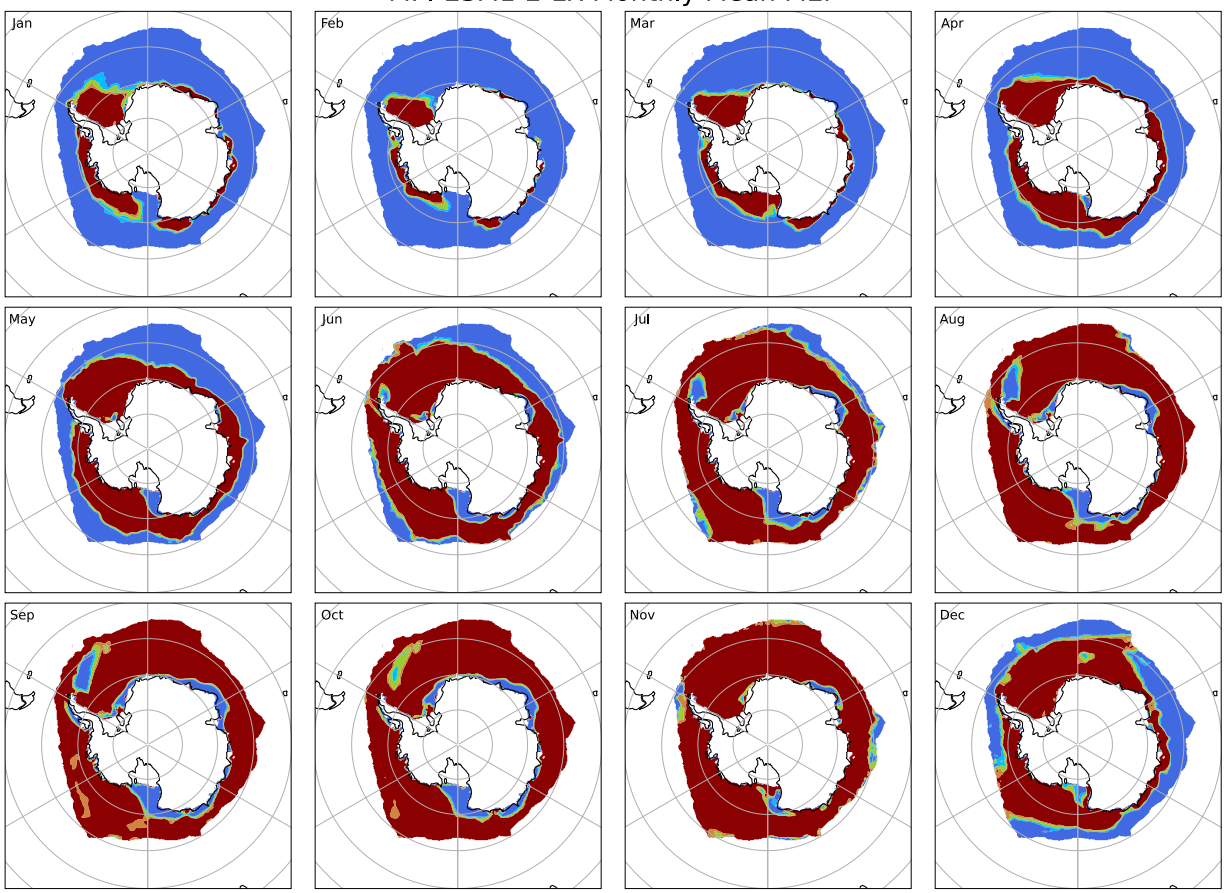


Figure B-64 MPI-ESM1-2-LR monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

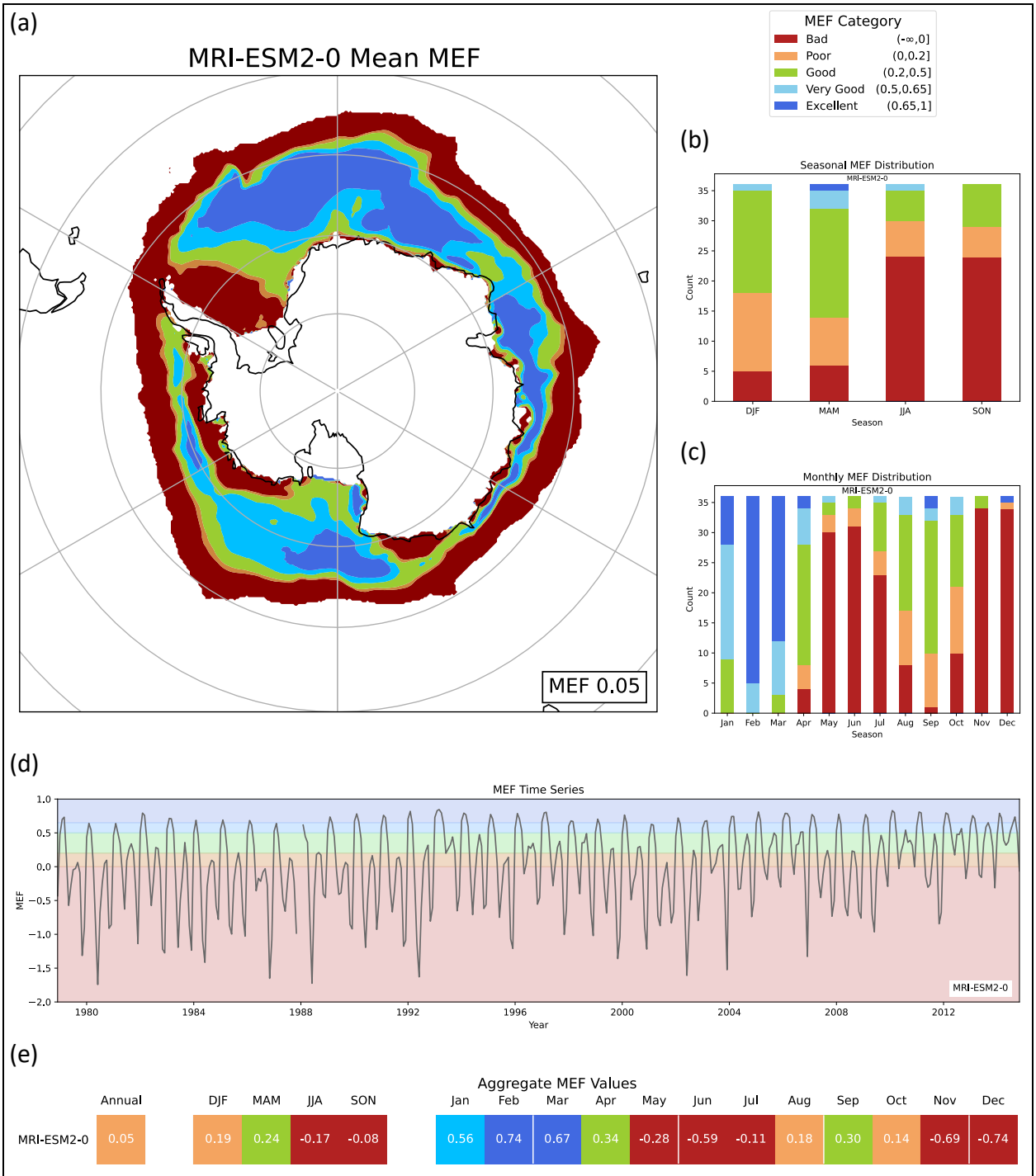


Figure B-65 MRI-ESM2-0 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

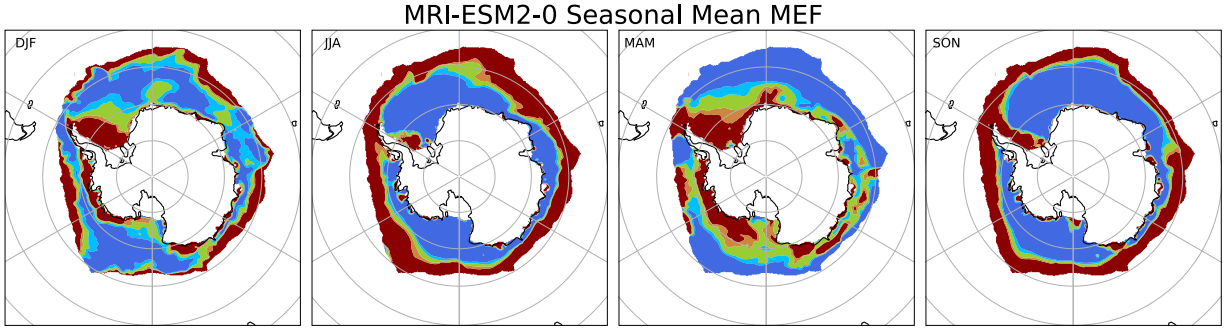


Figure B-66 MRI-ESM2-0 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

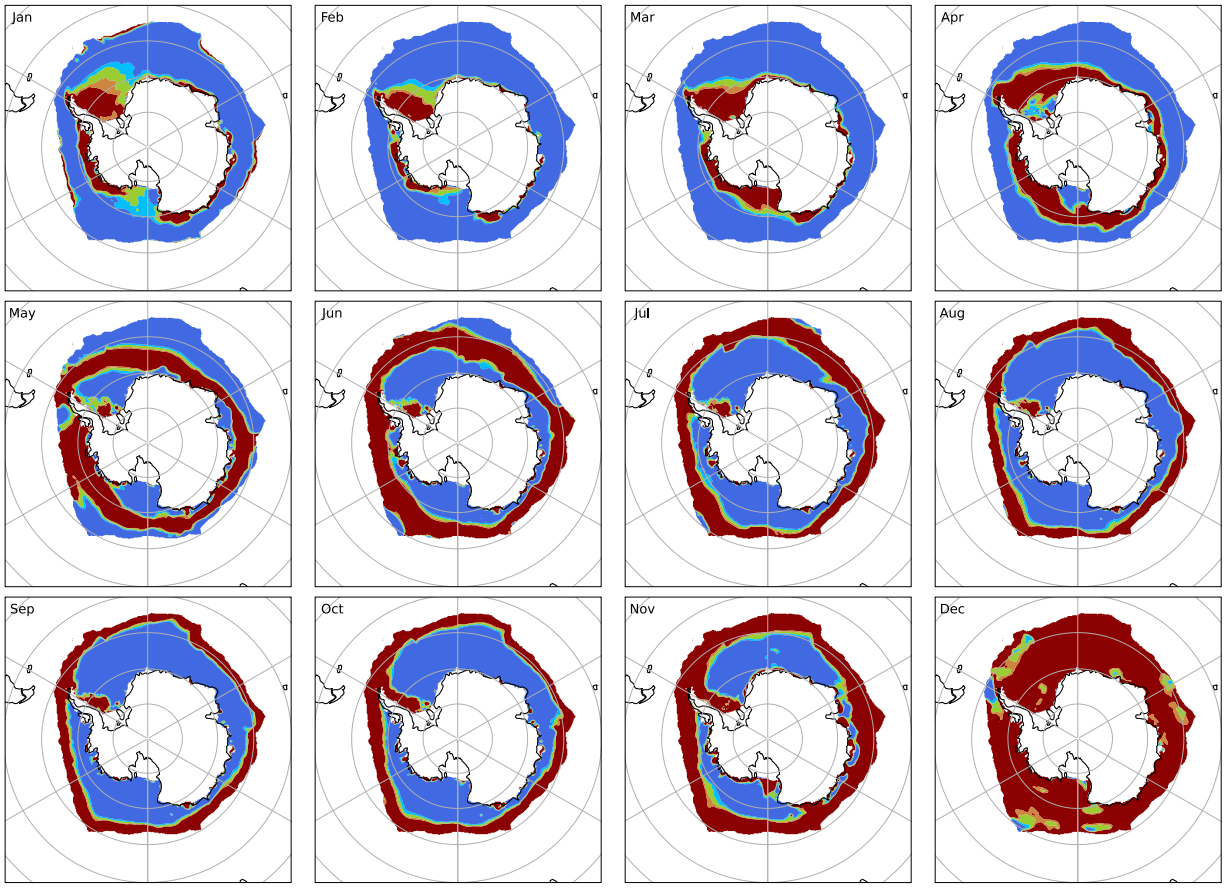


Figure B-67 MRI-ESM2-0 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

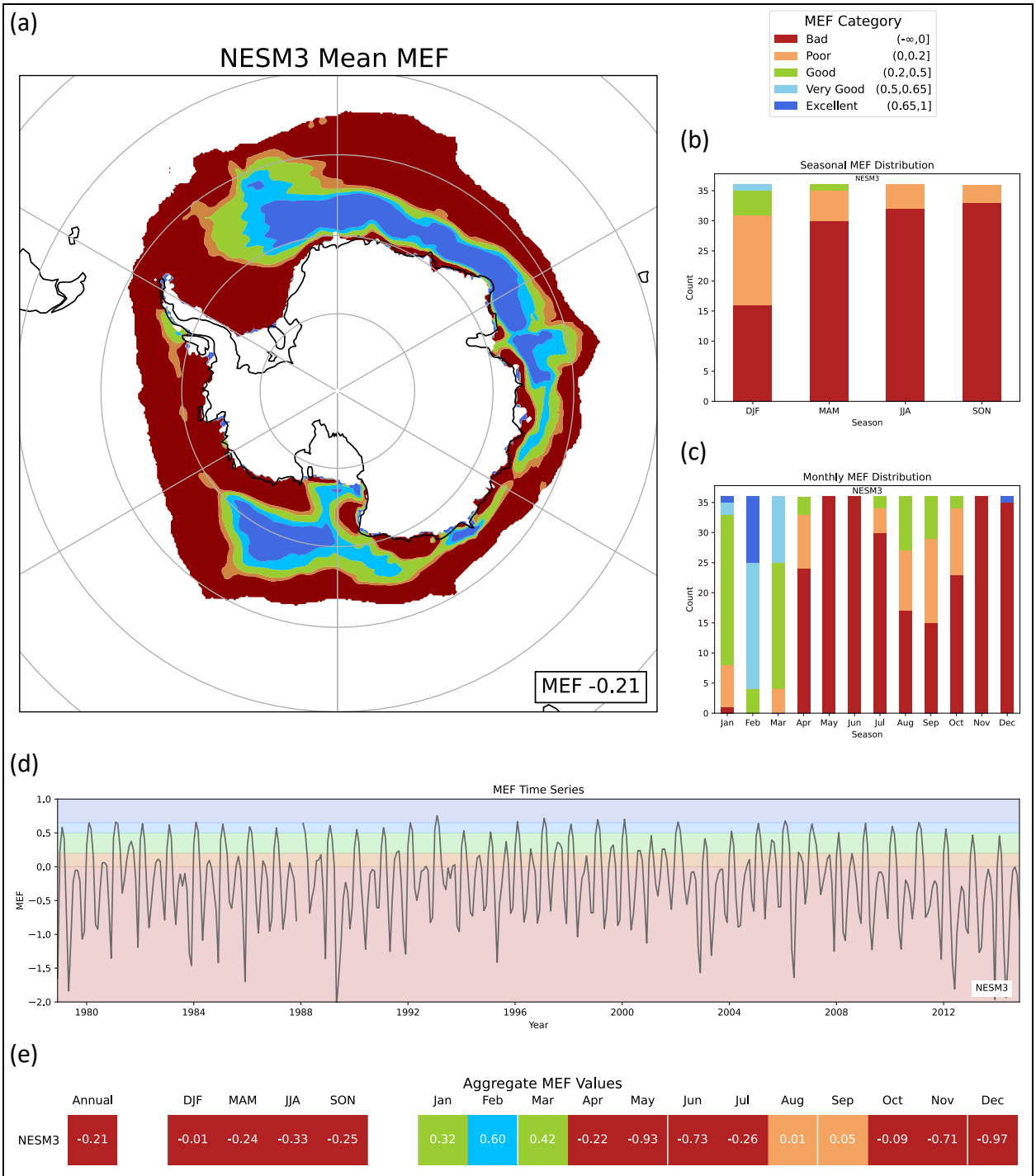


Figure B-68 NESM3 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

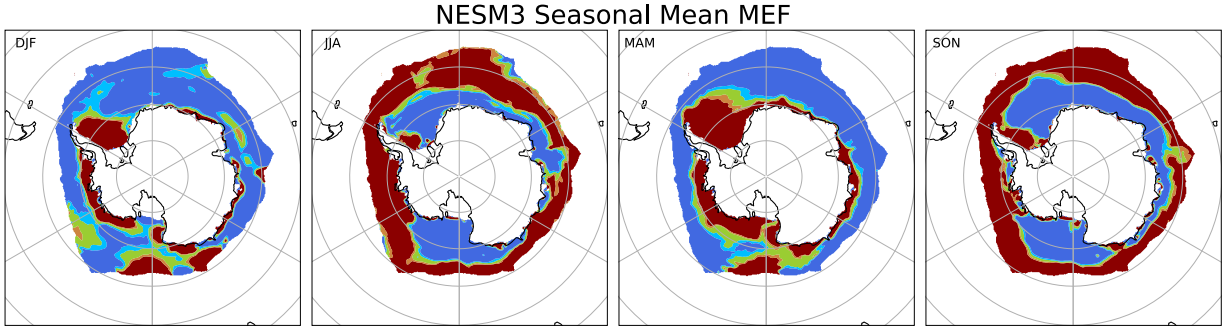


Figure B-69 NESM3 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

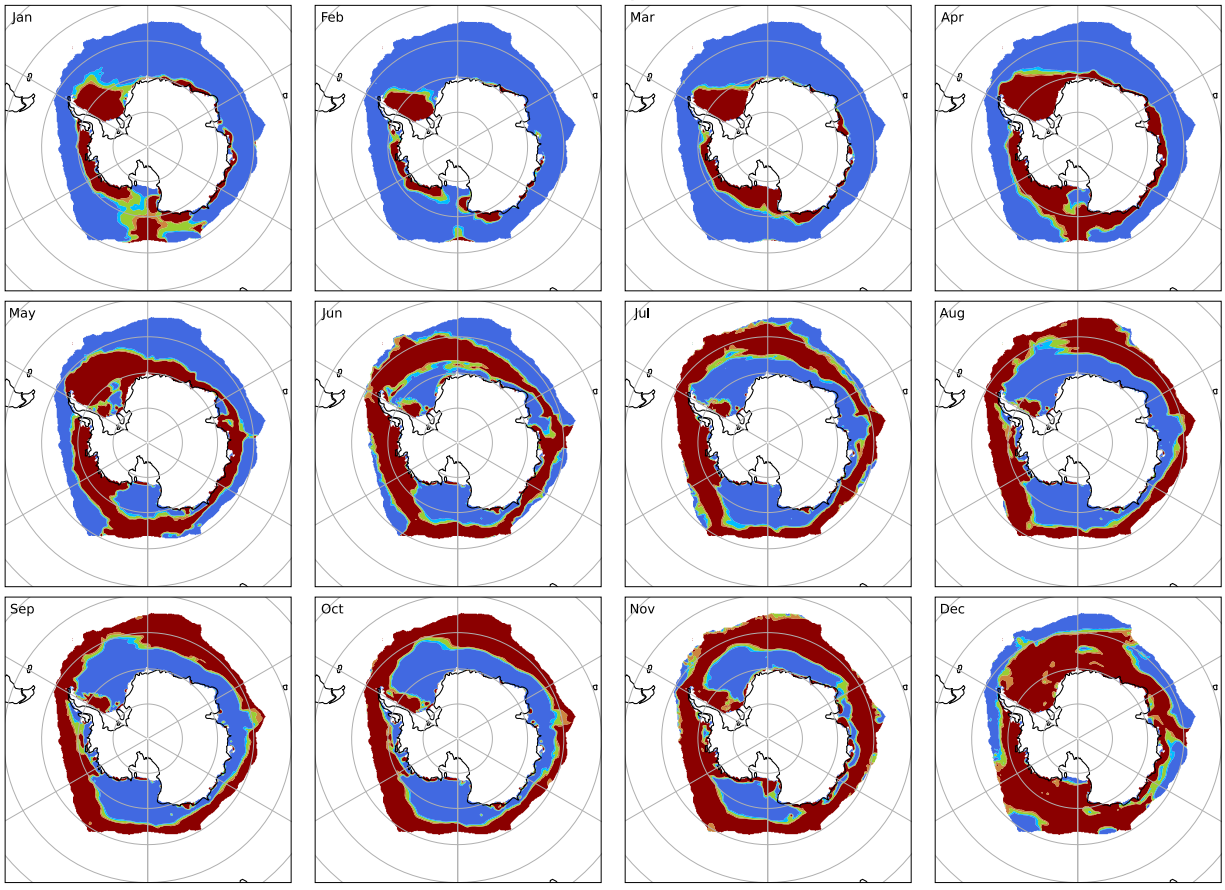


Figure B-70 NESM3 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

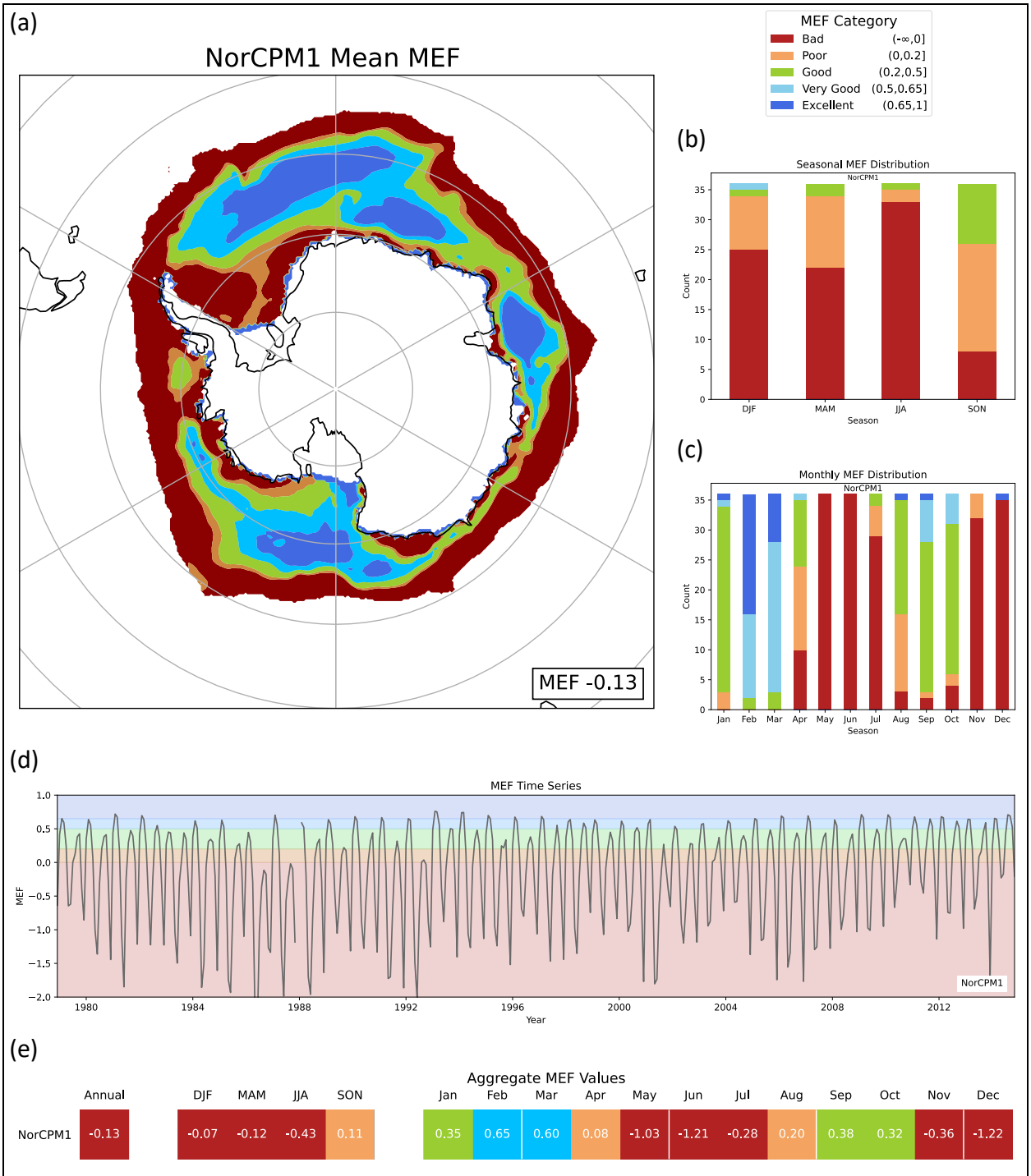


Figure B-71 NorCPM1 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

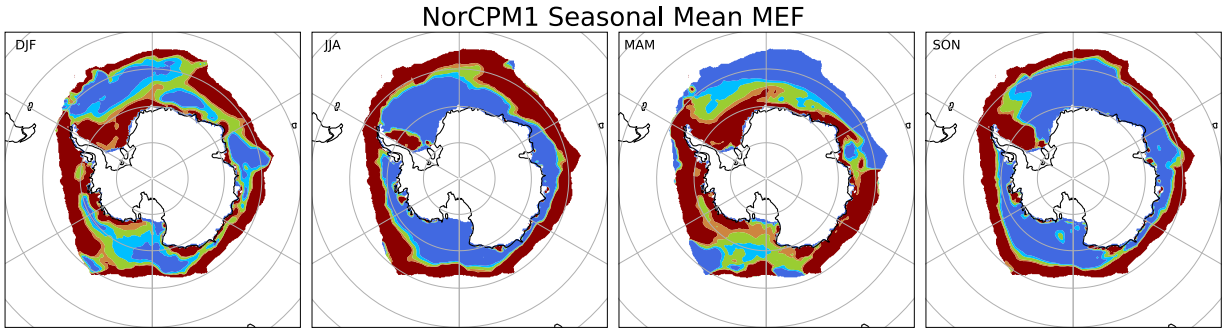


Figure B-72 NorCPM1 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

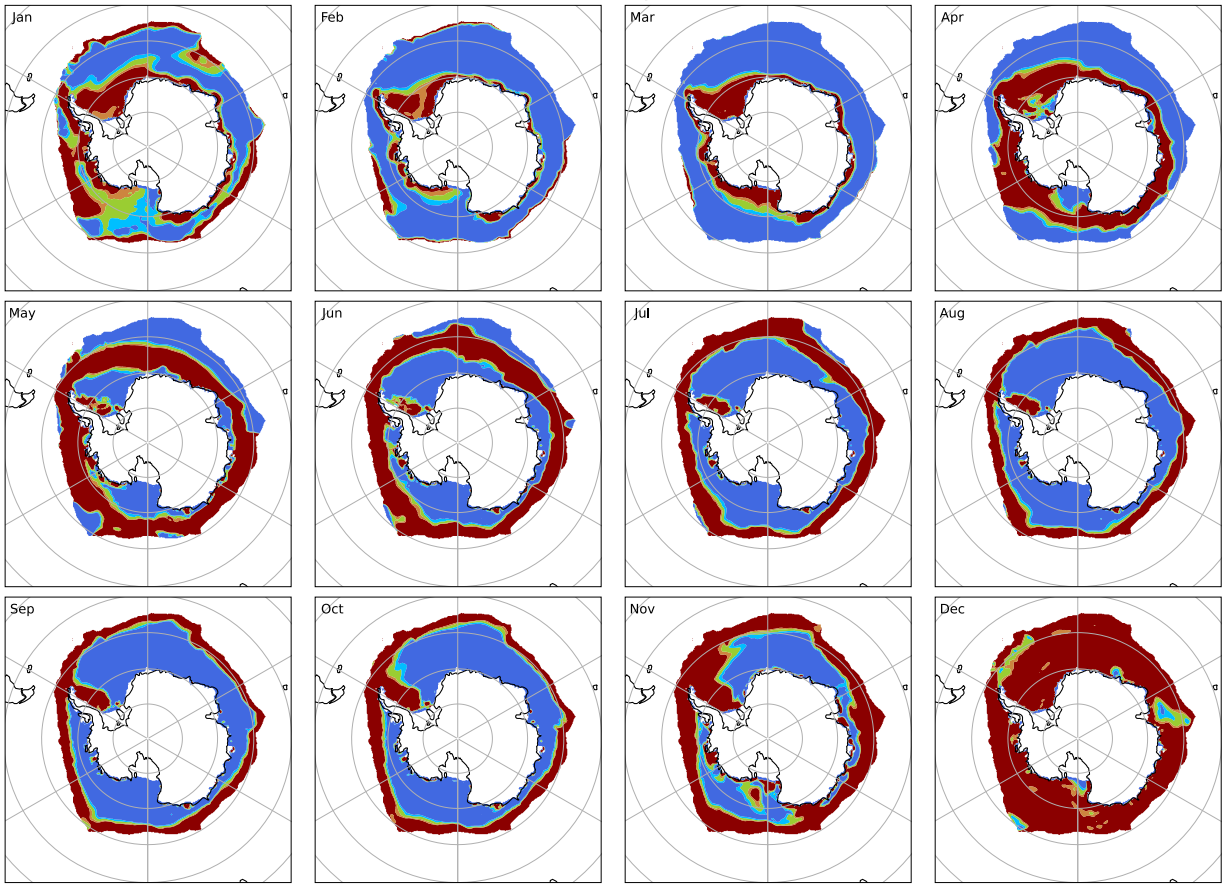


Figure B-73 NorCPM1 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

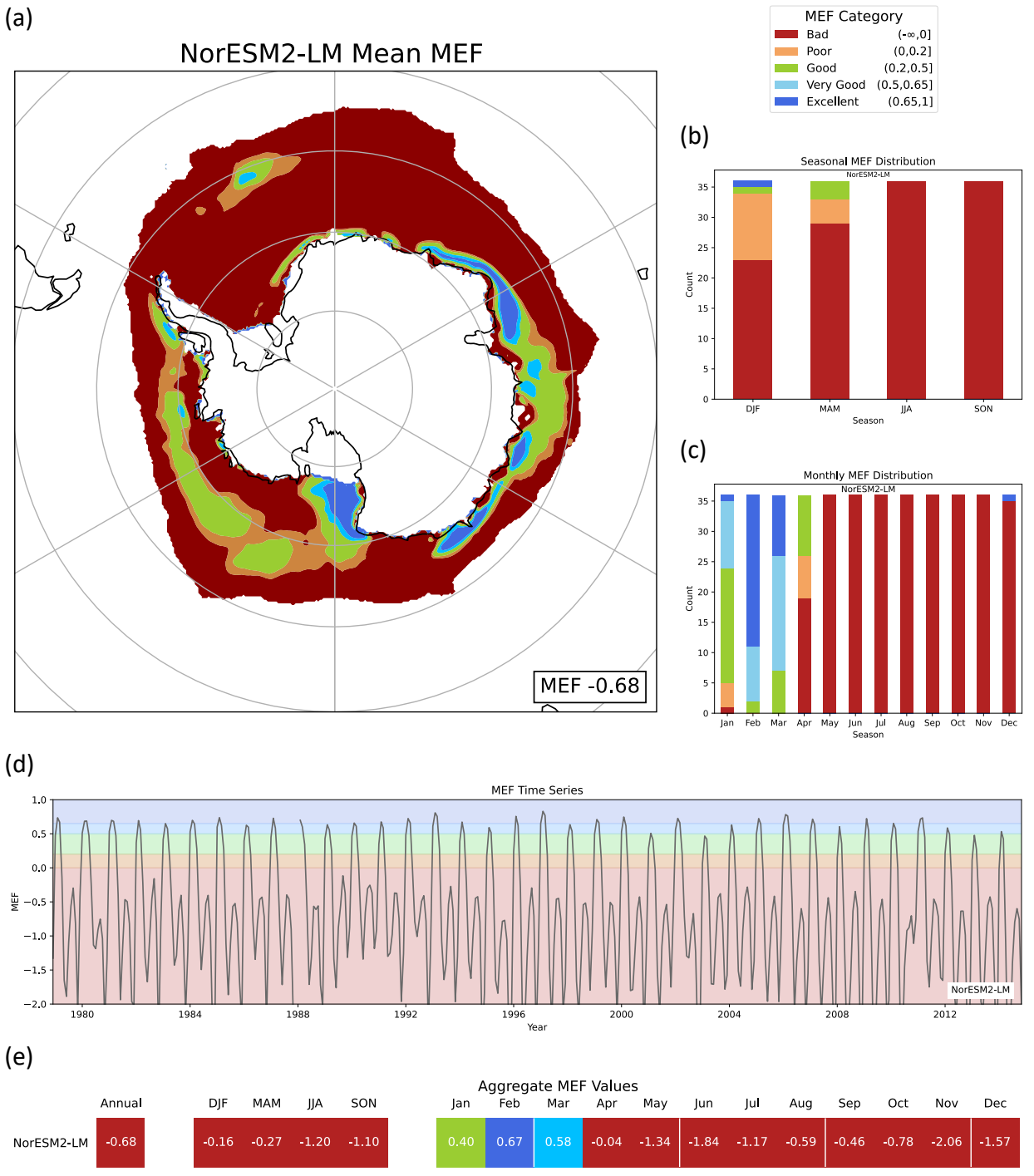


Figure B-74 NorESM2-LM MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

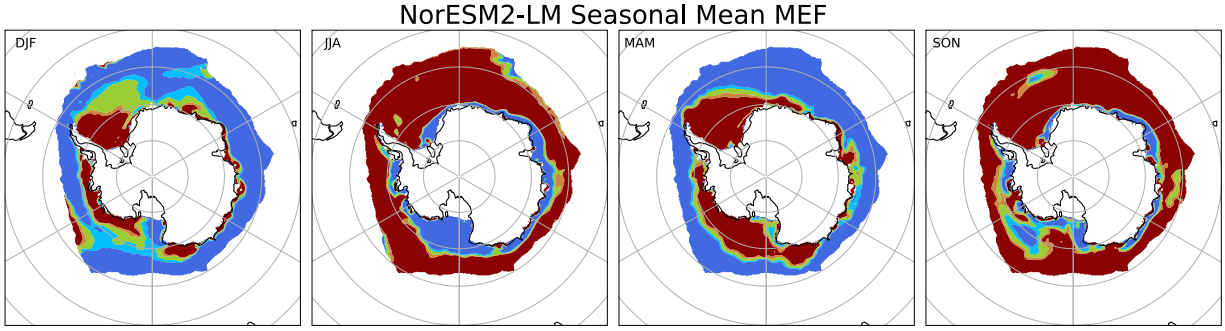


Figure B-75 NorESM2-LM seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

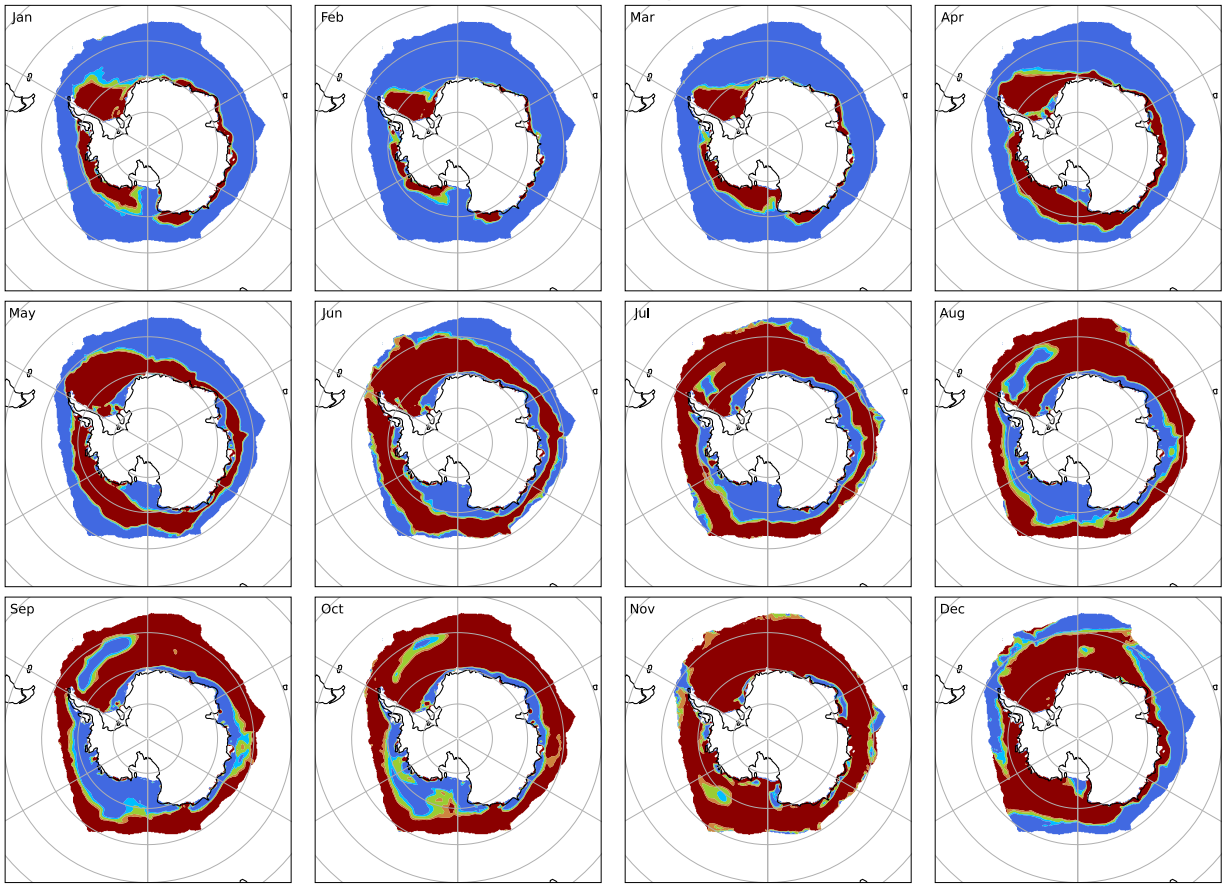


Figure B-76 NorESM2-LM monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

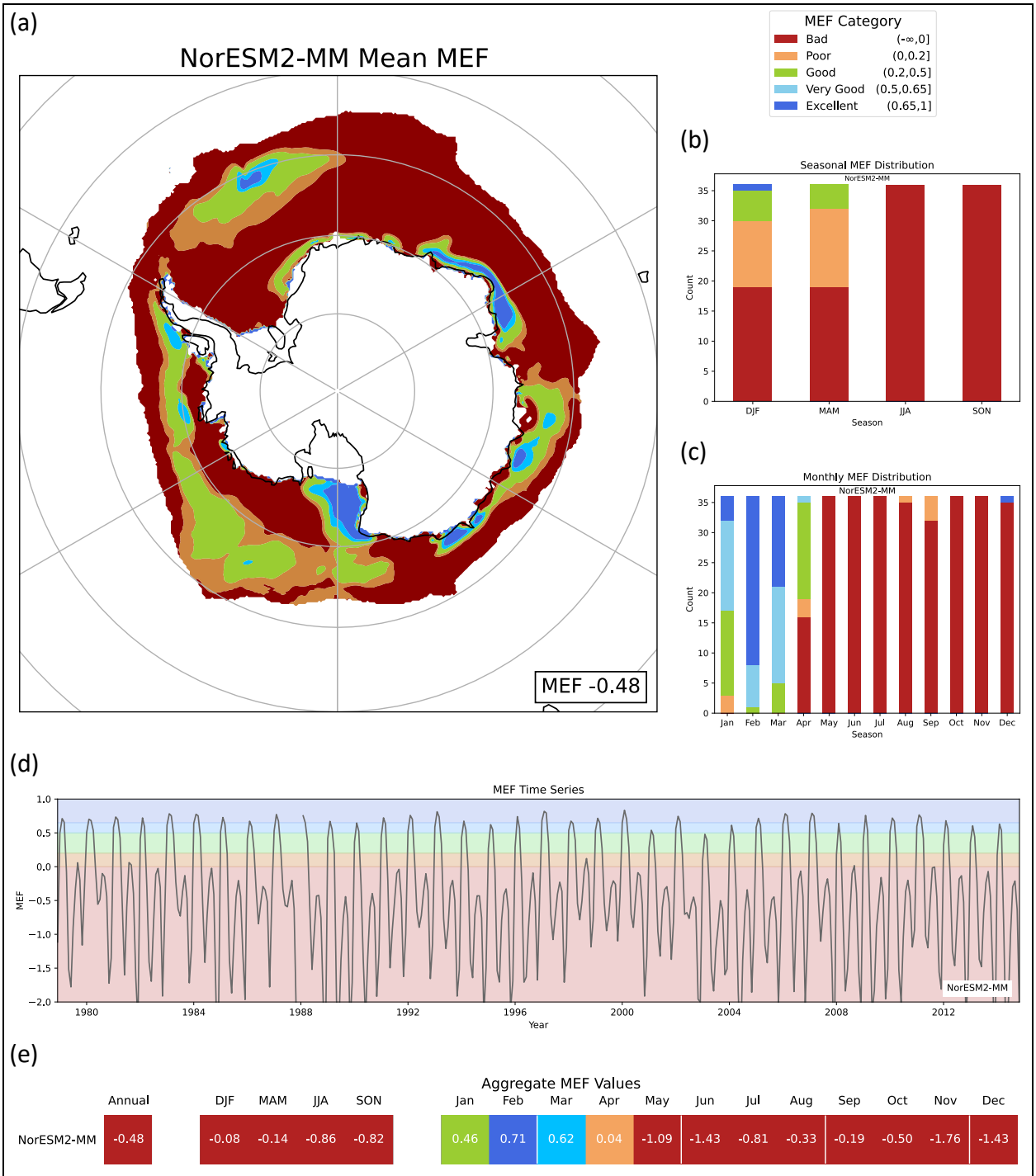


Figure B-77 NorESM2-MM MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

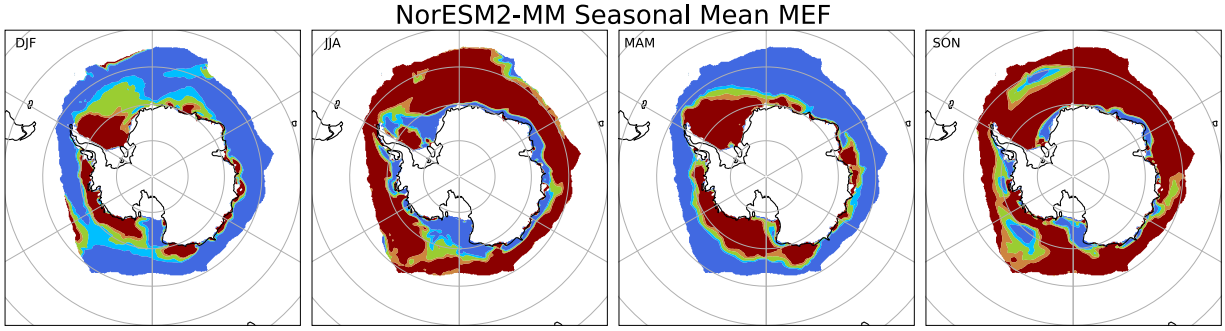


Figure B-78 NorESM2-MM seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

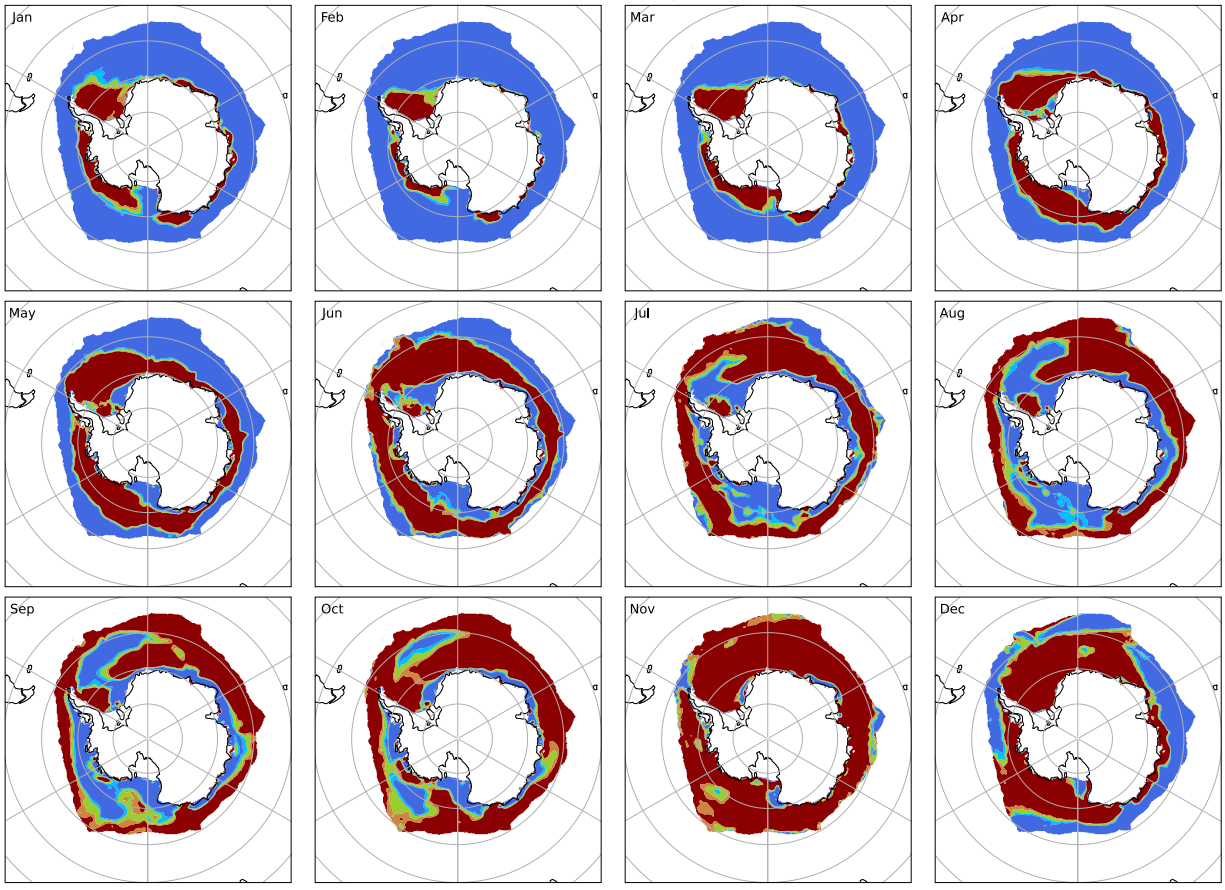


Figure B-79 NorESM2-MM monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

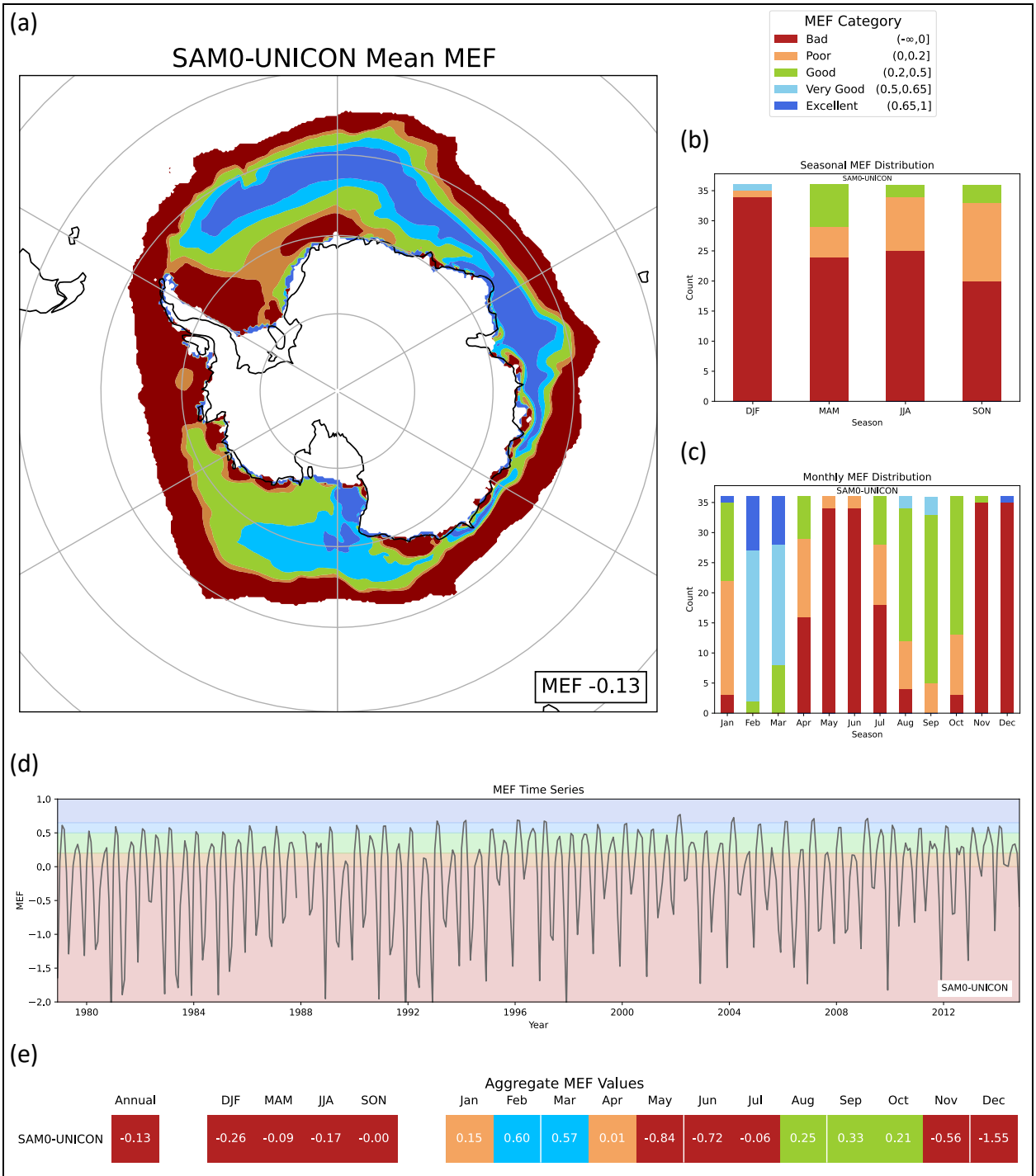


Figure B-80 SAM0-UNICON MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

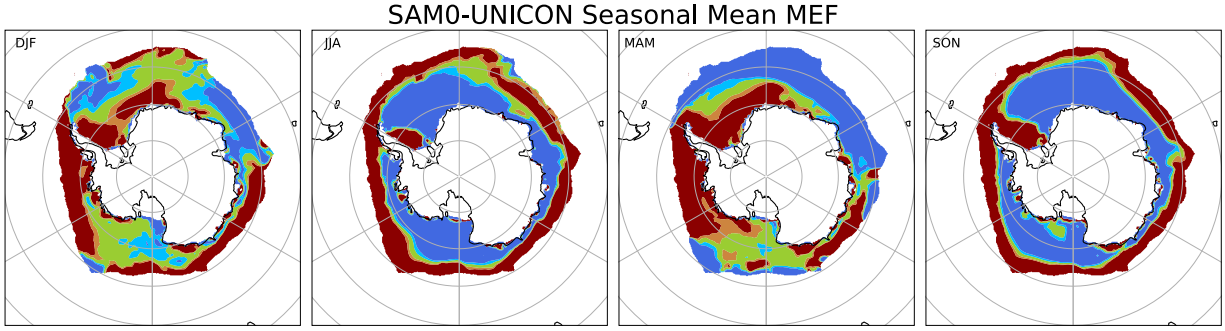


Figure B-81 SAM0-UNICON seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

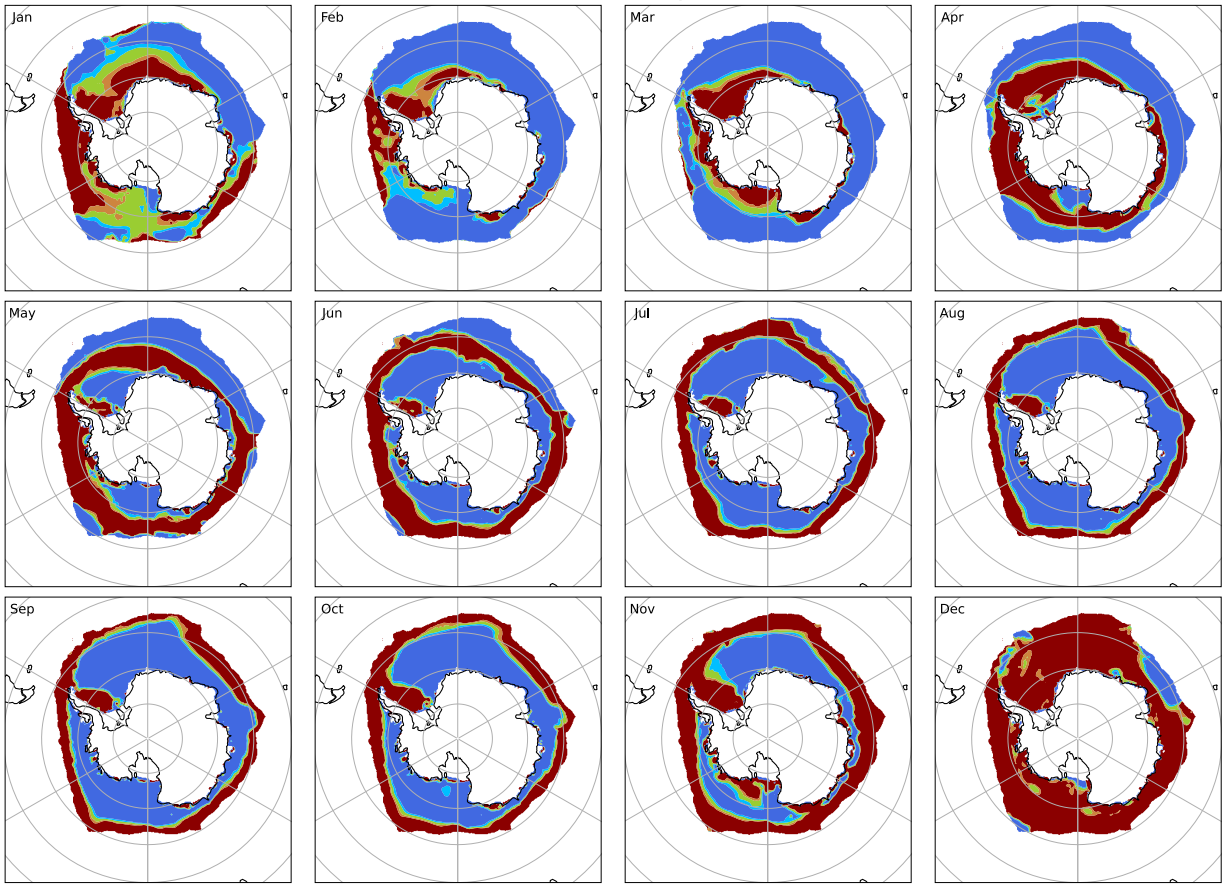


Figure B-82 SAM0-UNICON monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

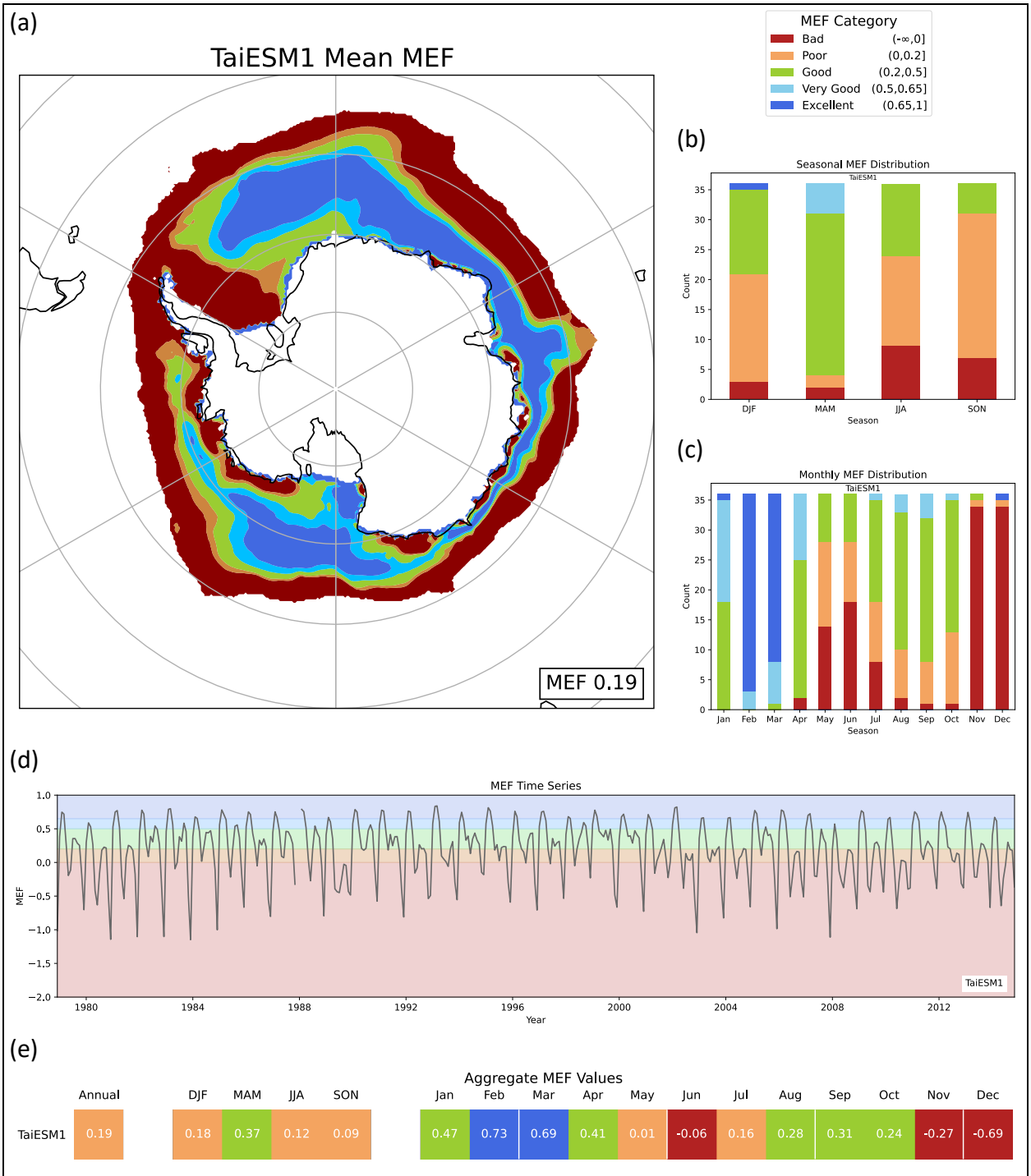


Figure B-83 TaiESM1 MEF metric plots

(a) Cumulative spatial mean MEF values for each valid grid point. Cumulative MEF index value displayed in the bottom right corner. (b) Seasonal temporal mean MEF distribution. (c) Monthly temporal mean MEF distribution. (d) Monthly temporal MEF value time series. (e) Cumulative MEF performance index values for annual, seasonal, and monthly categories.

Values colour mapped to categorical performance colours defined in Figure 3-1.

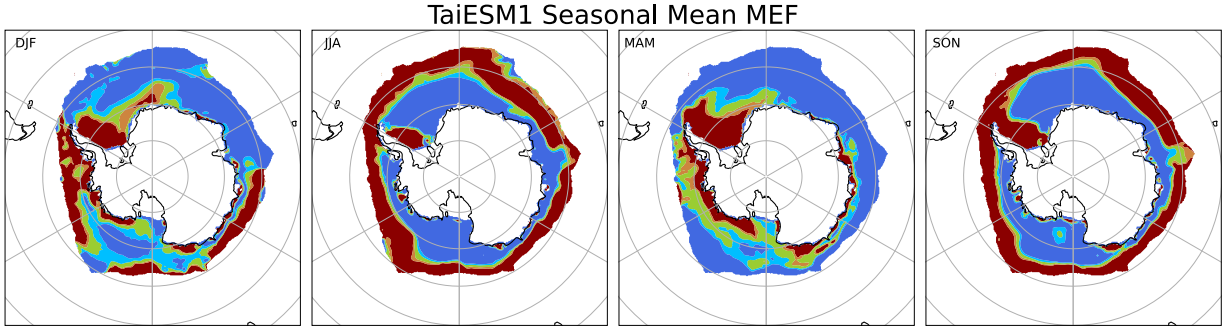


Figure B-84 TaiESM1 seasonal temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

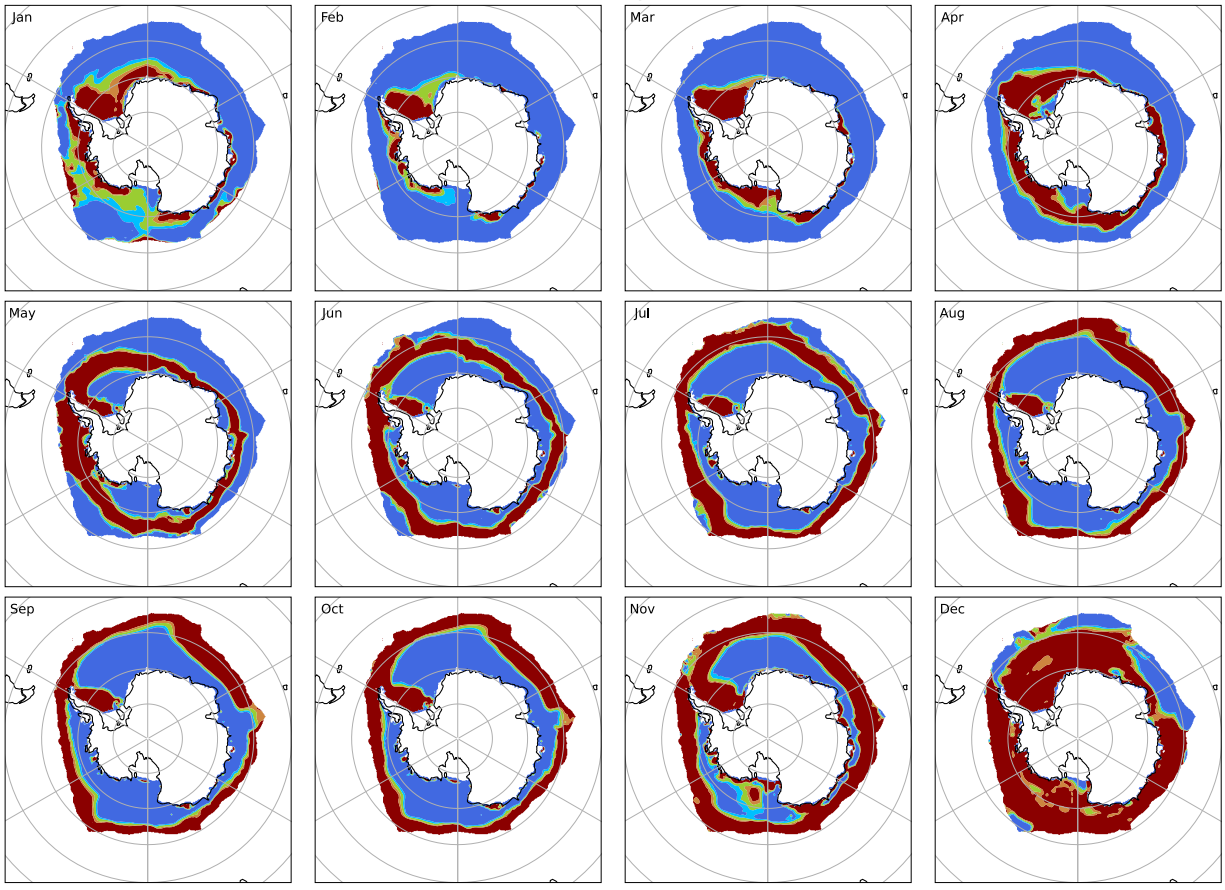


Figure B-85 TaiESM1 monthly temporal mean MEF plots
Values colour mapped to categorical performance colours defined in Figure 3-1.

Appendix C
CMIP6 Models Sensitivity Analysis

This Appendix contains the cumulative temporal mean metric plots and aggregated metric heatmaps for the sensitivity analysis of all 24 CMIP6 models generated during initial testing of the toolset. Plots are provided in alphabetical order in line with Table 1. Each figure contains the following plots: (a) Cumulative annual temporal mean MEF for the default configuration and all five sensitivity analysis configurations, and (b) Aggregated mean MEF annual, seasonal, and monthly values for all configurations. Colour contours and heat mapped values follow categories defined in Figure 3-1.

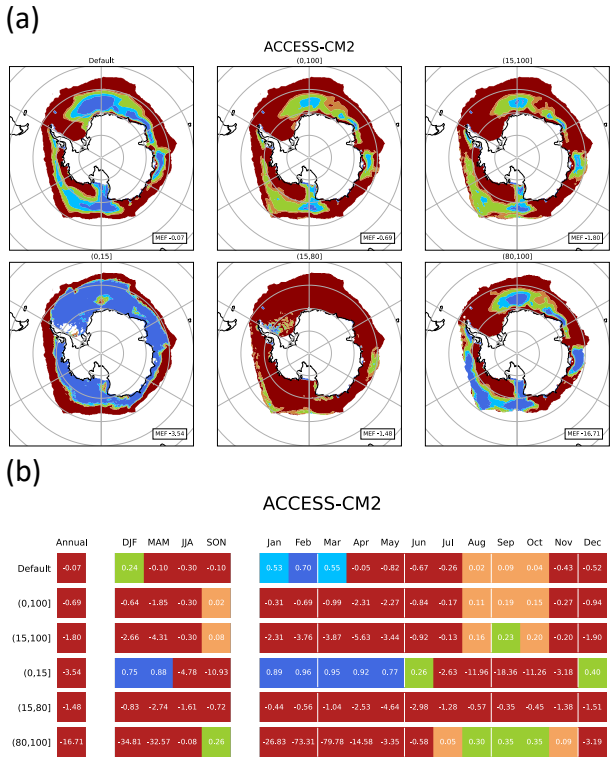


Figure C-86 ACCESS-CM2 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

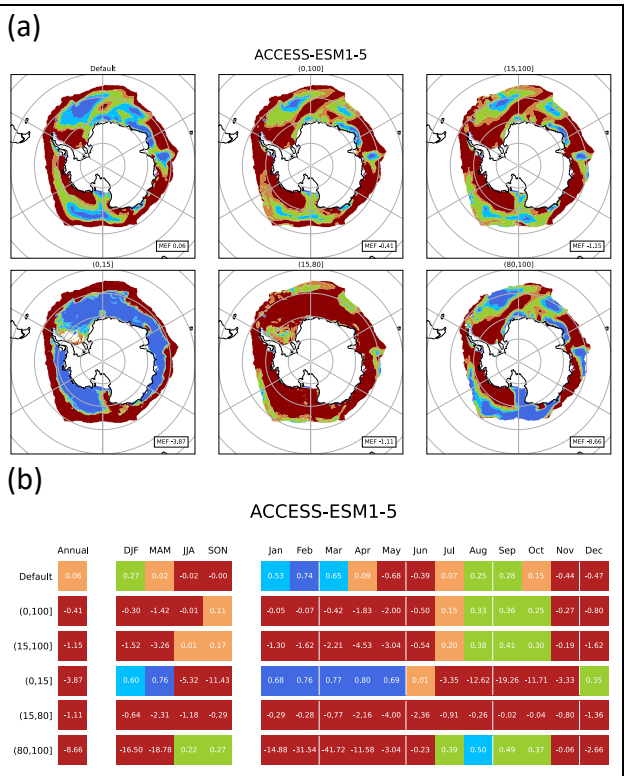
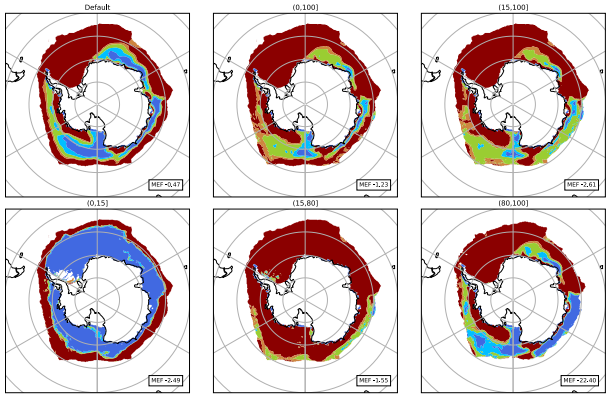


Figure C-87 ACCESS-ESM1-5 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)

BCC-CSM2-MR



(b)

BCC-CSM2-MR

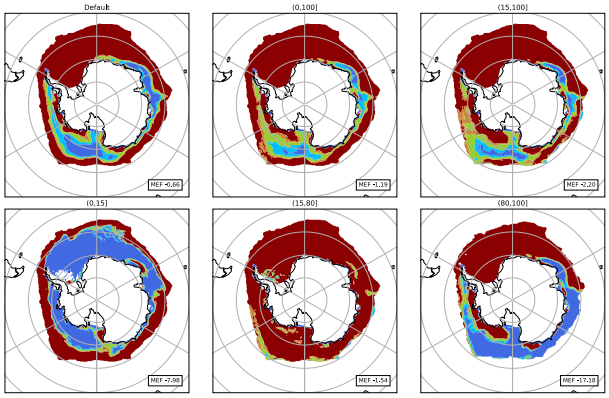
	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	-0.47	-0.04	-0.20	-0.88	-0.74	0.36	0.66	0.51	-0.12	-1.00	-1.33	-0.83	-0.48	-0.41	-0.52	-1.30	-1.16
(0,100]	-1.23	-1.27	-2.10	-0.89	-0.66	-0.86	-0.95	-1.18	-2.54	-2.59	-1.52	-0.74	-0.39	-0.32	-0.43	-1.23	-2.02
(15,100]	-2.61	-4.16	-4.80	-0.90	-0.63	-3.89	-4.51	-4.36	-6.14	-3.89	-1.63	-0.71	-0.36	-0.29	-0.39	-1.21	-4.06
(0,15]	-2.49	0.89	0.87	-4.33	-7.34	0.97	0.98	0.97	0.94	0.71	-0.34	-3.50	-9.15	-11.99	-7.84	-2.18	0.74
(15,80]	-1.55	-1.39	-2.41	-1.47	-0.93	-0.93	-0.76	-1.14	-2.23	-3.86	-2.75	-1.13	-0.52	-0.38	-0.57	-1.85	-2.49
(80,100]	-22.40	-48.27	-40.28	-0.90	-0.63	-44.33	-90.94	-98.48	-17.91	-4.44	-1.57	-0.72	-0.39	-0.33	-0.42	-1.14	-8.31

Figure C-88 BCC-CSM2-MR sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)

BCC-ESM1



(b)

BCC-ESM1

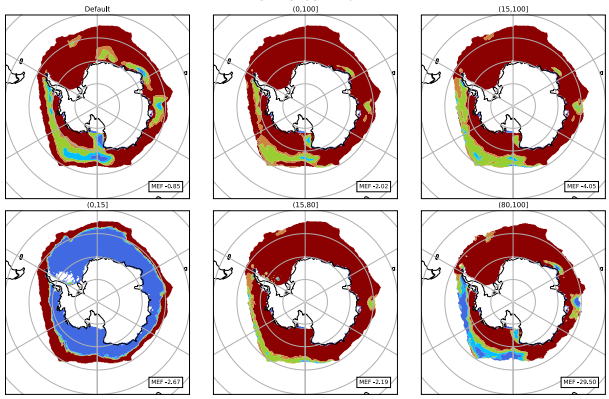
	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	-0.66	-0.08	-0.18	-1.27	-1.10	0.43	0.68	0.56	-0.01	-1.11	-1.75	-1.24	-0.81	-0.68	-0.84	-1.79	-1.37
(0,100]	-1.19	-1.09	-1.71	-1.07	-0.89	-0.53	-0.71	-0.81	-1.94	-2.38	-1.67	-0.94	-0.61	-0.51	-0.64	-1.52	-2.06
(15,100]	-2.20	-3.40	-3.67	-0.99	-0.79	-2.84	-3.69	-3.21	-4.53	-3.27	-1.63	-0.82	-0.51	-0.42	-0.54	-1.40	-3.65
(0,15]	-7.98	0.66	0.49	-12.69	-20.23	0.88	0.91	0.83	0.62	0.03	-2.05	-10.10	-25.94	-33.47	-20.30	-6.93	0.18
(15,80]	-1.54	-1.25	-2.09	-1.70	-1.12	-0.51	-0.53	-0.76	-1.76	-3.76	-3.02	-1.40	-0.69	-0.46	-0.65	-2.26	-2.72
(80,100]	-17.18	-38.02	-29.37	-0.95	-0.79	-33.09	-73.55	-72.41	-12.24	-3.44	-1.49	-0.81	-0.55	-0.49	-0.59	-1.28	-6.40

Figure C-89 BCC-ESM1 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)

CAMS-CSM1-0



(b)

CAMS-CSM1-0

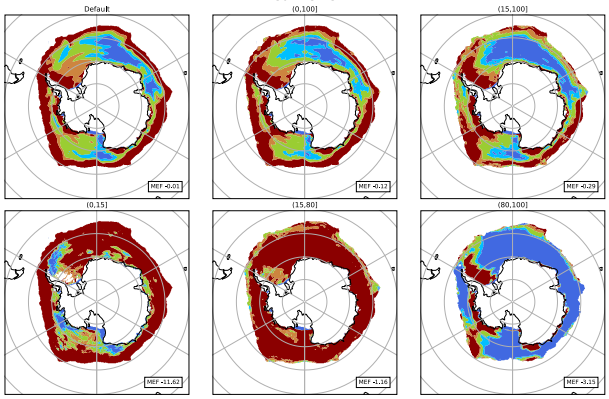
	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	-0.85	-0.59	-1.00	-1.05	-0.76	0.21	0.62	0.39	-0.82	-2.56	-2.10	-0.82	-0.23	-0.06	-0.26	-1.96	-2.65
(0,100]	-2.02	-2.23	-4.04	-1.13	-0.68	-1.31	-1.16	-1.75	-4.82	-5.55	-2.49	-0.76	-0.14	0.03	-0.16	-1.90	-4.26
(15,100]	-4.05	-6.11	-8.27	-1.20	-0.66	-5.15	-5.12	-5.82	-10.85	-8.15	-2.75	-0.75	-0.10	0.07	-0.12	-1.93	-8.09
(0,15]	-2.67	0.96	0.96	-3.80	-8.75	0.98	0.99	0.98	0.98	0.93	0.59	-1.83	-10.16	-16.25	-9.16	-0.83	0.91
(15,80]	-2.19	-2.23	-3.68	-1.77	-1.07	-1.30	-0.87	-1.57	-3.42	-6.06	-3.56	-1.26	-0.50	-0.27	-0.49	-2.45	-4.54
(80,100]	-29.50	-59.92	-56.79	-1.21	-0.63	-57.92	-103.18	-127.48	-32.92	-9.97	-2.84	-0.74	-0.06	0.12	-0.07	-1.95	-17.44

Figure C-90 CAMS-CSM1-0 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)

CanESM5



(b)

CanESM5

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	-0.01	-0.15	0.01	0.00	0.08	0.27	0.56	0.44	0.08	-0.50	-0.41	0.07	0.36	0.42	0.29	-0.46	-1.30
(0,100]	-0.12	-0.48	-0.57	0.25	0.31	-0.12	-0.03	-0.17	-0.67	-0.87	-0.13	0.35	0.53	0.56	0.46	-0.10	-1.30
(15,100]	-0.29	-1.18	-0.91	0.45	0.47	-0.90	-1.05	-0.83	-1.17	-0.73	0.14	0.53	0.67	0.67	0.59	0.14	-1.61
(0,15]	-11.62	-0.04	-0.41	-18.21	-27.59	0.33	0.48	0.25	-0.22	-1.27	-3.41	-13.95	-37.26	-45.86	-26.55	-10.35	-0.96
(15,80]	-1.16	-0.82	-2.00	-1.19	-0.63	-0.17	-0.20	-0.59	-1.67	-3.75	-2.39	-0.90	-0.26	-0.11	-0.23	-1.55	-2.10
(80,100]	-3.15	-10.56	-3.73	0.78	0.76	-10.34	-20.11	-10.90	-0.63	0.34	0.66	0.82	0.88	0.86	0.81	0.62	-0.96

Figure C-91 CanESM5 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

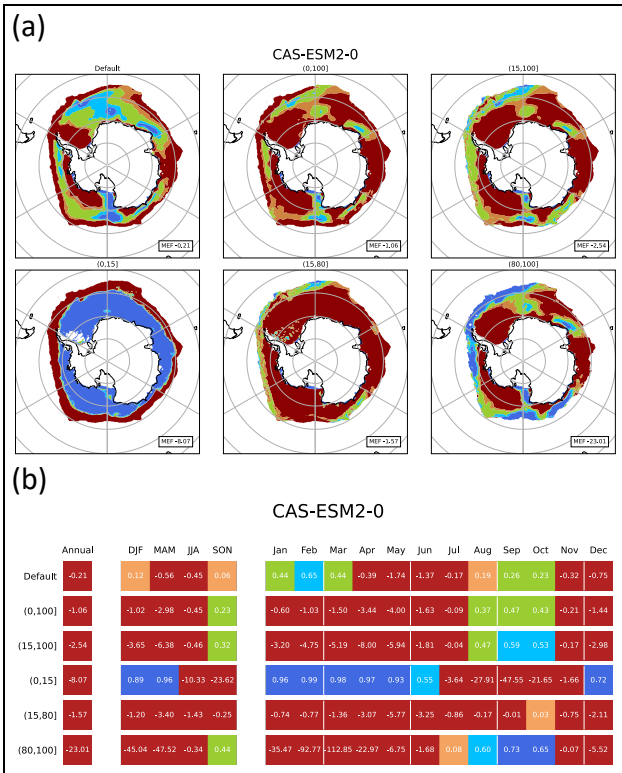


Figure C-92 CAS-ESM2-0 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

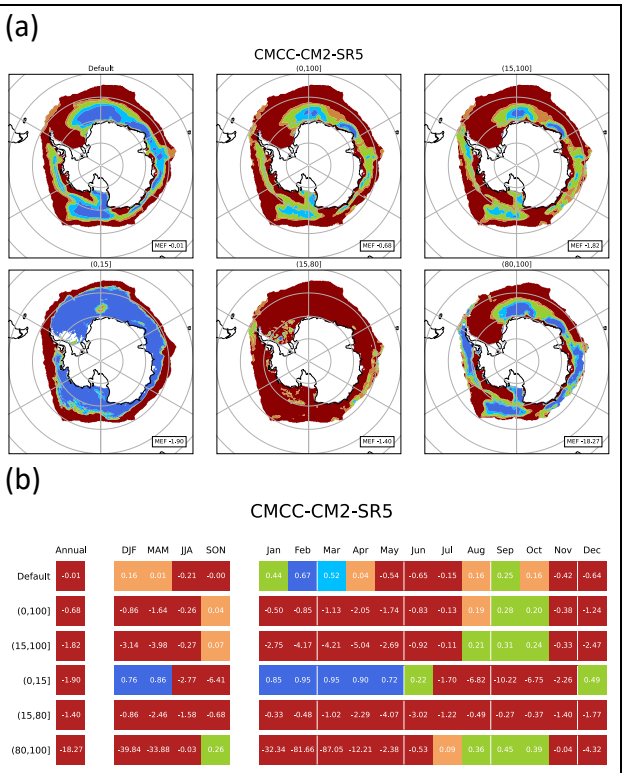


Figure C-93 CMCC-CM2-SR5 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

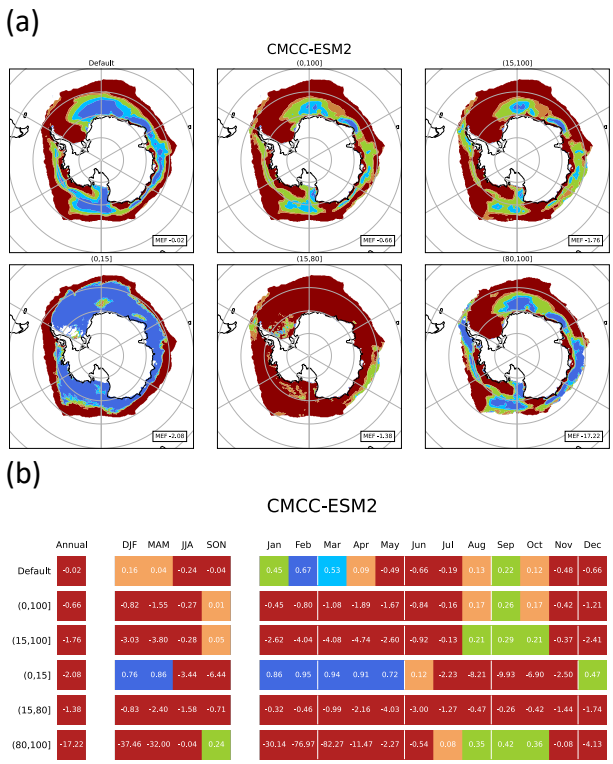


Figure C-94 CMCC-ESM2 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

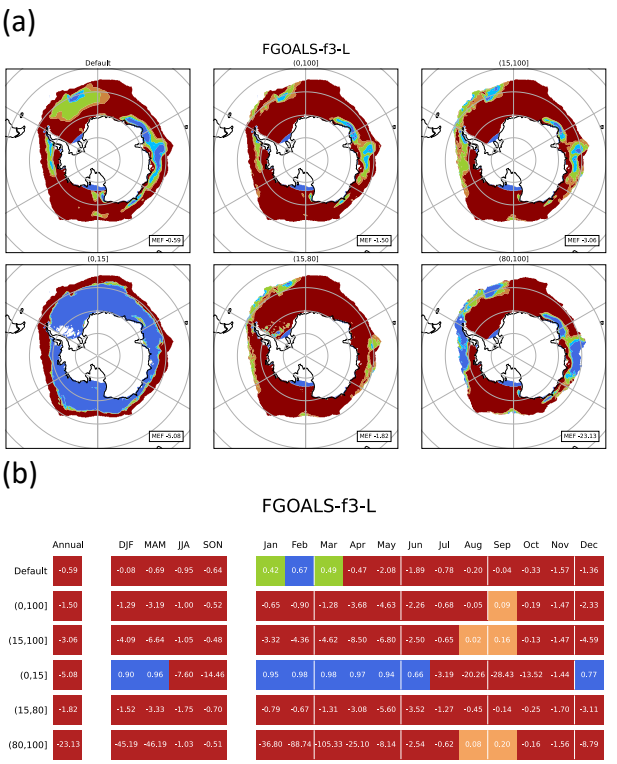


Figure C-95 FGOALS-f3-L sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

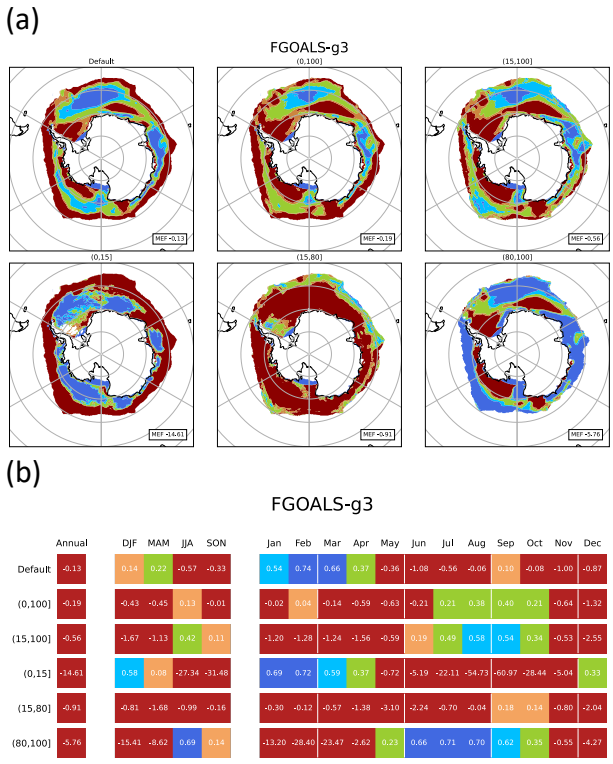


Figure C-96 FGOALS-g3 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

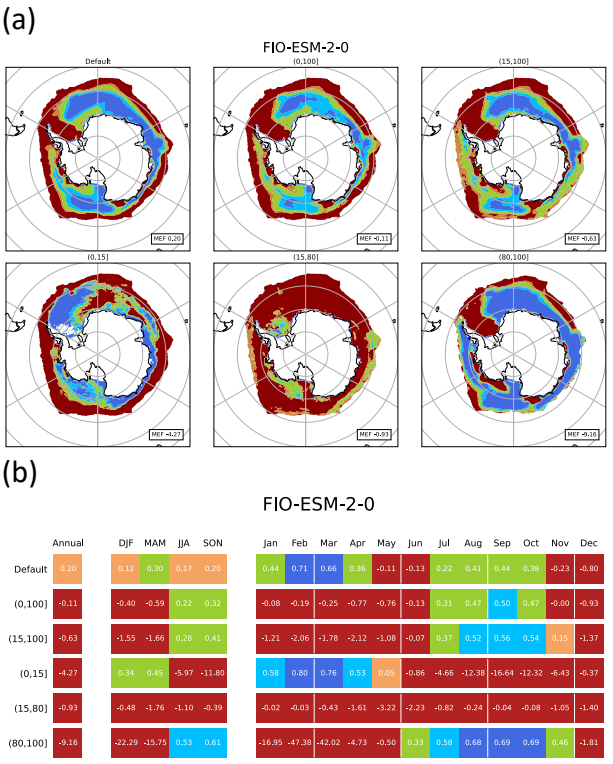


Figure C-97 FIO-ESM-2-0 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

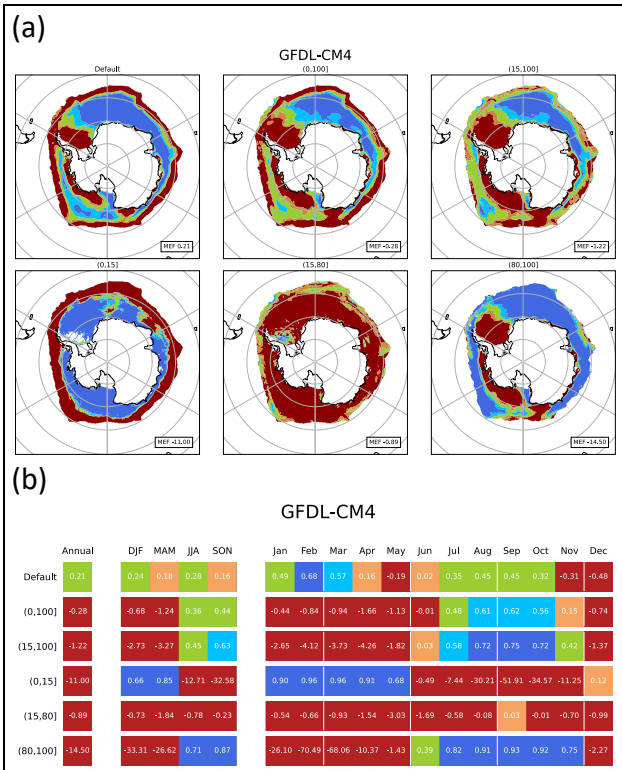


Figure C-98 GFDL-CM4 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

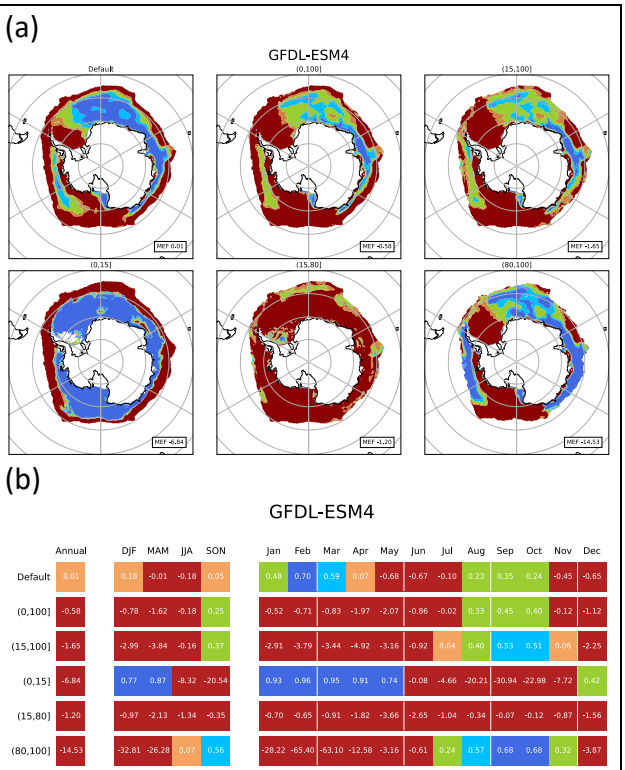
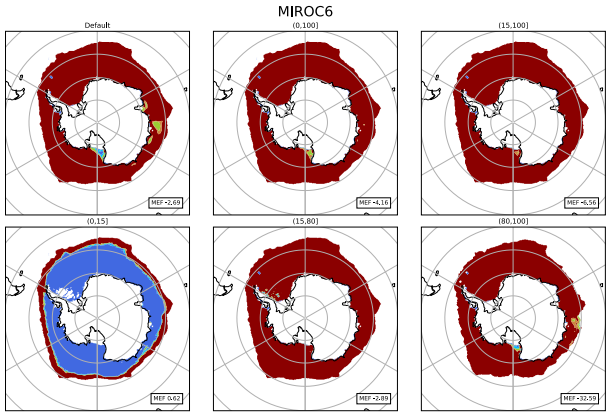


Figure C-99 GFDL-ESM4 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

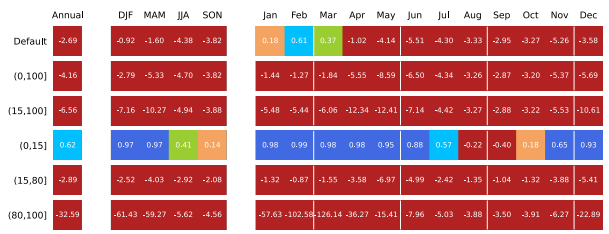
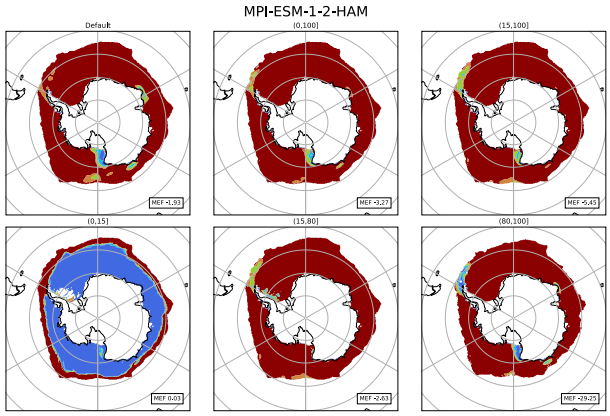


Figure C-100 MIROC6 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

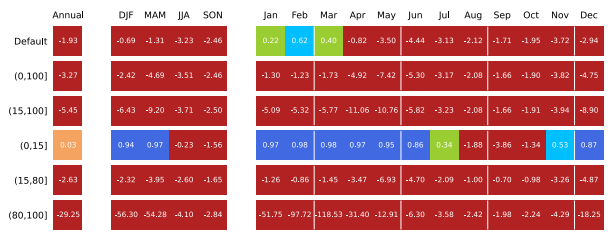
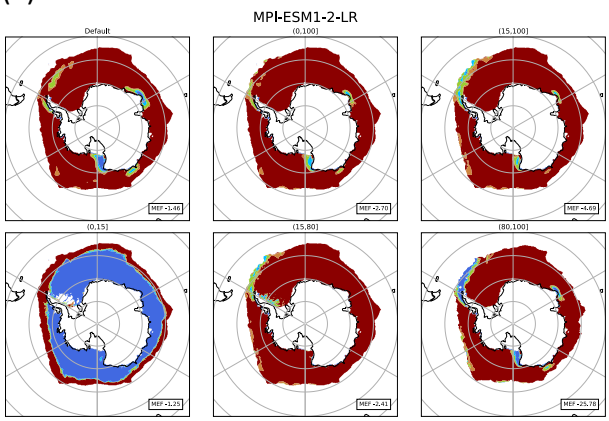


Figure C-101 MPI-ESM-1-2-HAM sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

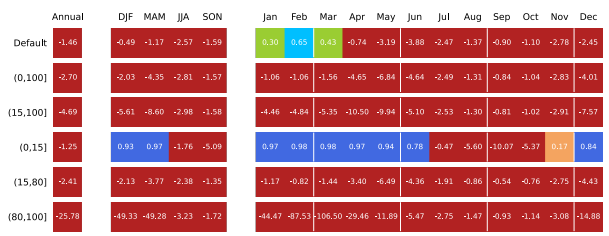
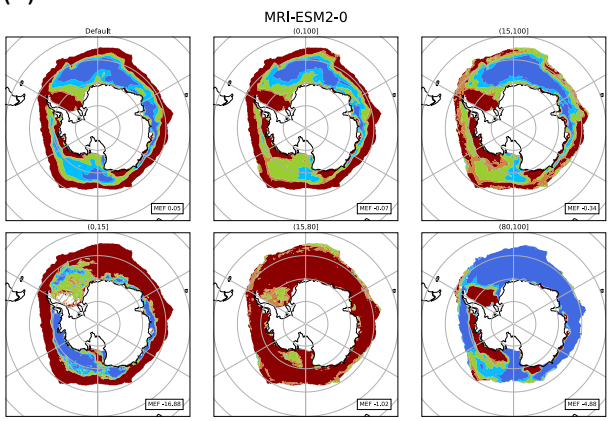


Figure C-102 MPI-ESM1-2-LR sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

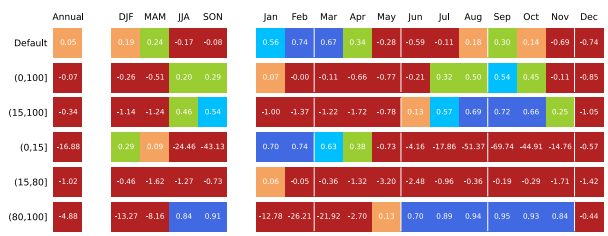
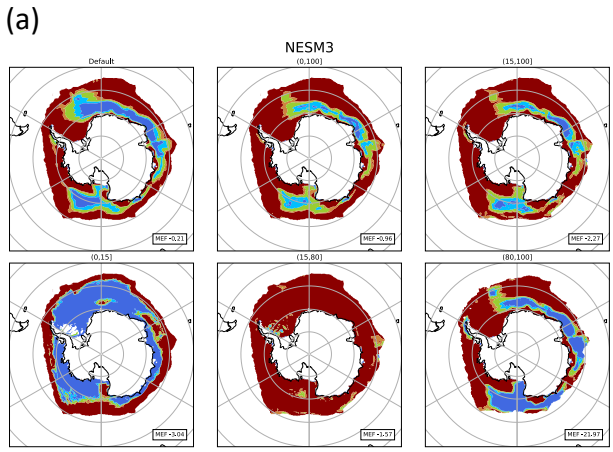


Figure C-103 MRI-ESM2-0 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.



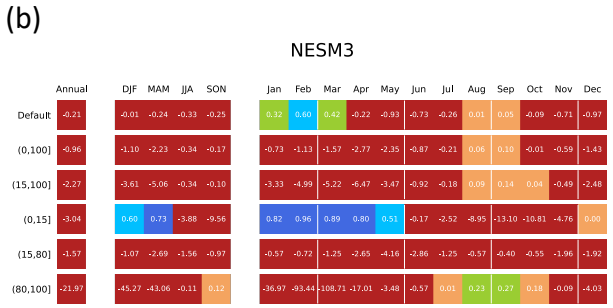
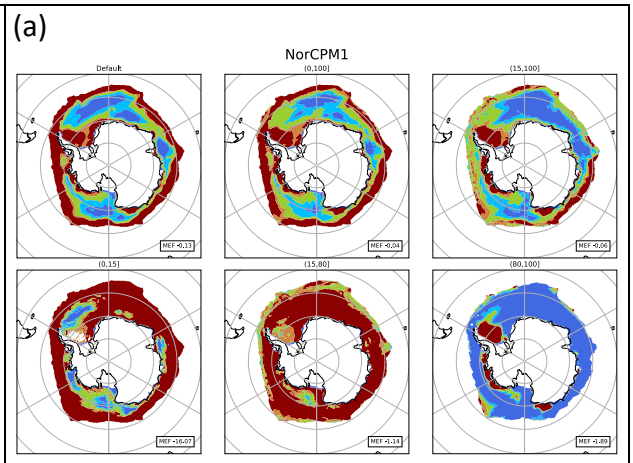


Figure C-104 NESM3 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.



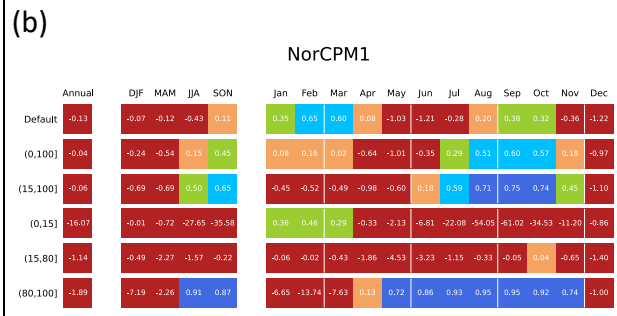
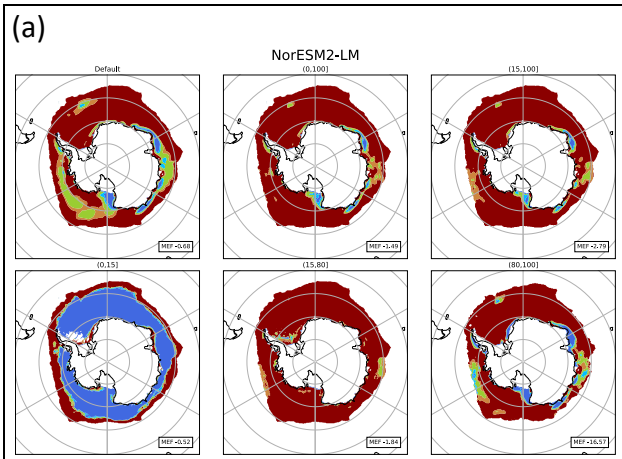


Figure C-105 NorCPM1 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.



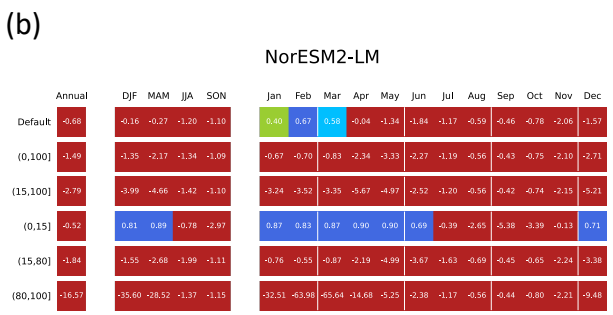
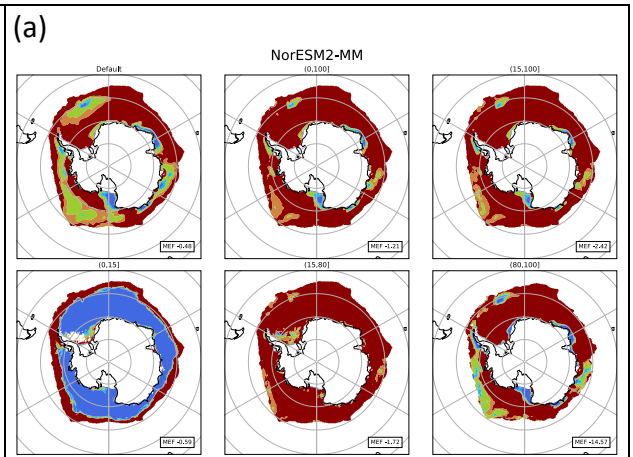


Figure C-106 NorESM2-LM sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.



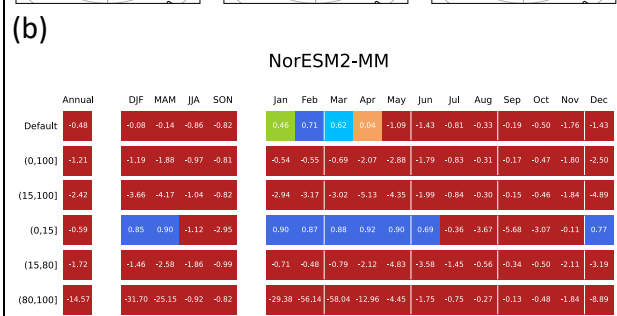
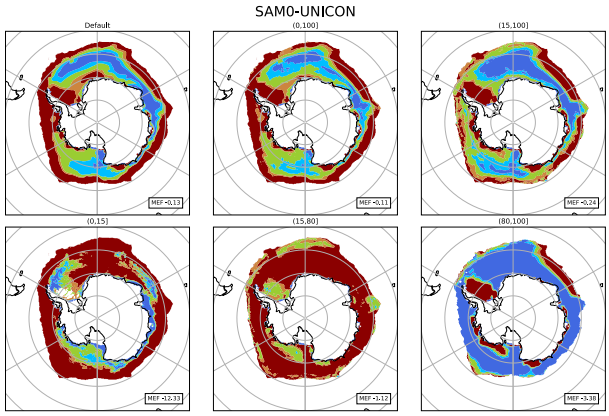


Figure C-107 NorESM2-MM sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

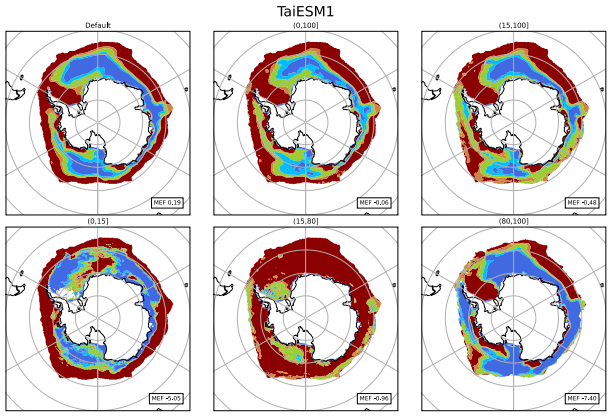
SAM0-UNICON

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	-0.13	-0.26	-0.09	-0.17	-0.00	0.15	0.60	0.57	0.01	-0.84	-0.72	-0.06	0.25	0.33	0.21	-0.56	-1.55
(0,100]	-0.11	-0.36	-0.61	0.21	0.32	-0.08	0.12	-0.04	-0.78	-0.99	-0.16	0.33	0.47	0.51	0.44	0.02	-1.13
(15,100]	-0.24	-0.92	-0.98	0.44	0.50	-0.69	-0.80	-0.83	-1.37	-0.73	0.18	0.53	0.61	0.63	0.58	0.29	-1.27
(0,15]	-12.33	-0.05	-0.49	-18.67	-29.88	0.25	0.58	0.46	-0.24	-1.69	-4.21	-14.92	-36.87	-48.87	-29.23	-11.56	-1.01
(15,80]	-1.12	-0.56	-2.21	-1.29	-0.43	-0.08	0.05	-0.44	-1.94	-4.26	-2.64	-0.93	-0.29	-0.08	-0.14	-1.06	-1.66
(80,100]	-3.38	-10.30	-4.86	0.79	0.74	-9.28	-20.37	-14.10	-0.96	0.47	0.74	0.81	0.81	0.80	0.76	0.66	-0.95

Figure C-108 SAM0-UNICON sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.

(a)



(b)

TaiESM1

	Annual	DJF	MAM	JJA	SON	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Default	0.19	0.18	0.37	0.12	0.09	0.47	0.73	0.69	0.41	0.01	-0.06	0.16	0.28	0.31	0.24	-0.27	-0.69
(0,100]	-0.06	-0.24	-0.44	0.19	0.24	0.05	-0.01	-0.12	-0.63	-0.56	-0.03	0.25	0.35	0.39	0.35	-0.03	-0.78
(15,100]	-0.48	-1.15	-1.39	0.26	0.33	-0.82	-1.49	-1.48	-1.86	-0.83	0.04	0.33	0.42	0.45	0.42	0.13	-1.12
(0,15]	-5.05	0.32	0.51	-7.08	-13.87	0.54	0.76	0.78	0.58	0.15	-0.86	-5.31	-15.06	-20.37	-14.76	-6.47	-0.35
(15,80]	-0.96	-0.44	-1.68	-1.18	-0.52	0.01	-0.00	-0.42	-1.55	-3.08	-2.19	-0.98	-0.37	-0.15	-0.22	-1.20	-1.34
(80,100]	-7.40	-17.44	-13.39	0.52	0.54	-12.88	-37.73	-36.15	-3.82	-0.20	0.45	0.56	0.57	0.57	0.57	0.47	-1.14

Figure C-109 TaiESM1 sensitivity analysis plots

Values colour mapped to categorical performance colours defined in Figure 3-1.