The updated Multi-Model Large Ensemble Archive and the Climate Variability Diagnostics Package: New tools for the study of climate variability and change

Nicola Maher[1,2], Adam S. Phillips[3], Clara Deser[3], Robert C. Jnglin Wills[4], Flavio Lehner[3,5,6], John Fasullo[3], Julie M. Caron[3], Lukas Brunner[7], Urs Beyerle[4], and Jemma Jeffree[1]

[1]Australian National University, Canberra, Australia
[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, USA
[3]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, United States
[4]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[5]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, United States
[6]Polar Bears International, Bozeman, MT, United States
[7]Affiliation: Research Unit Sustainability and Climate Risk, Center for Earth System Research and Sustainability (CEN), University of Hamburg, Hamburg, Germany

Correspondence: Nicola Maher (nicola.maher@anu.edu.au)

Abstract.

Observations can be considered as one realisation of the climate system that we live in. To provide a fair comparison of climate models with observations, one must use multiple realisations or *ensemble members* from a single model and assess where the observations sit within the ensemble spread. Single model initial-condition large ensembles (LEs) are valuable tools for such an evaluation. Here, we present the new multi-model large ensemble archive (MMLEAv2) which has been extended to include 18 models and 16 two-dimensional variables. Data in this archive has been remapped to a common 2.5 x 2.5 degree grid for ease of inter-model comparison. We additionally introduce the newly updated Climate Variability Diagnostics Package version 6 (CVDPv6), which is designed specifically for use with LEs. The CVDPv6 computes and displays the major modes of climate variability as well as long-term trends and climatologies in models and observations based on a variety of fields. This tool creates plots of both individual ensemble members, and the ensemble mean of each LE including observational rank plots, pattern correlations and root mean square difference metrics displayed in both graphical and statistical output that is saved to a data repository. By applying the CVDPv6 to the MMLEAv2 we highlight its use for model evaluation against observations and for model inter-comparisons. We demonstrate that for highly variable metrics a model might evaluate poorly or favourably compared to the single realisation the observations represent, depending on the chosen ensemble member. This behaviour emphasises that LEs provide a much fairer model evaluation than a single ensemble member, ensemble mean, or multi-model mean. By leveraging the combination of the CVDPv6 and MMLEAv2 presented in this paper we can determine which climate variables need a large ensemble for fair assessment against observations.

1 Introduction

Single model initial-condition large ensembles (LEs) are a powerful tool for understanding past, current and future climate (e.g. Deser et al., 2020; Maher et al., 2021a). A single LE allows for both the quantification and separation of the modelled forced response (response of any given variable to external forcing) and unforced internal variability, while the availability of multiple LEs additionally enables the assessment of model differences in both quantities (Deser et al., 2020; Lee et al., 2021; Maher et al., 2021b; Wood et al., 2021; Maher et al., 2023). LEs also facilitate a robust evaluation of individual climate models in comparison to observations, as they include a range of possible climate realisations (Goldenson et al., 2021; Suarez-Gutierrez et al., 2021; Labe and Barnes, 2022). Observations can be considered as one realisation of the climate system, and as such the interpretation of its comparison to a single historical simulation from a climate model, an ensemble mean, or multi-model mean is complicated. To this point, fairly evaluating projections in single runs of climate models, particularly for highly variable climate quantities against this single realisation of the real world is only possible by taking long time averages, to effectively smooth out natural climate variability and allowing for the assessment of the model's forced response. The advantage of using a LE is that we can additionally evaluate whether observations sit within the model's ensemble spread. This is a necessary, although not sufficient, condition for model evaluation that makes LEs invaluable tools for such evaluation. The value of LEs derives from their large sample size. In addition to providing a range of plausible outcomes arising from the superposition of forced response and internal variability, the value of LEs derives from the sheer volume of data that they provide, enabling robust statistics on climate variability. For example, 250 years of simulation (e.g. 1850-2100) in a LE of only 20 members yields 5000 years of data for analysis, while a typical pre-industrial control simulation (piControl) provides 500-2000 years of output.

While evaluating whether the observations sit within the model spread is crucial for model evaluation as it takes into account the fact that observations are only one possible realisation of the Earth system, understanding whether a model's forced response and internal variability are similar to observations is also vital. Quantifying the forced response and internal variability in observations is non-trivial and an active area of research. Ongoing efforts such as the “Forced Component Estimation Statistical Method Intercomparison Project” (ForceSMIP) (Wills et al., 2025), are expected to provide guidance on how best to extract forced response and internal variability from the observational record. LEs are critical to such projects. Quantifying these metrics in LEs themselves is more straightforward, but can be limited by ensemble size. The ensemble size needed to quantify any given metric is found to be dependent on the metric itself and the acceptable error on the quantification (Milinski et al., 2020; Lee et al., 2021; Planton et al., 2024). Progress has been made in the assessment of these quantities in combination in comparison to observations with studies such as Suarez-Gutierrez et al. (2021) using LEs and a rank-histogram method to evaluate whether LEs have a realistic forced response and internal variability of annual temperatures compared to the observed record. Such evaluations are becoming increasingly possible with the lengthening observational record and increasing forced trends (Simpson et al., 2025). Additionally, further understanding when the forced response (signal) is likely to emerge from the internal variability (noise) is vital to understanding the climate we observe (e.g. Hawkins and Sutton, 2012). This concept is important for climate communication and quantifying when we expect to see climate change signals appear in our observed record. This is another avenue where LEs are vital tools to further research in the area (e.g. Schlunegger et al., 2020). Finally, the recent discovery of the signal-to-noise paradox (Scaife and Smith, 2018) where an ensemble can better predict observations than their own members, is also an avenue where LEs can provide valuable insight into model behaviour (Weisheimer et al., 2024).

The Multi-Model Large Ensemble Archive (MMLEA; Deser et al., 2020) was the first compilation of many LEs. The overview paper has been cited over 650 times despite only being published in 2020 (4 years ago; data as of 22/10/24). This highlights the community's need for and the value of such an archive. In this paper, we present the MMLEAv2 archive that has been expanded beyond the original MMLEA to both include more models (18 compared to the original 7, largely those from the more recent CMIP6 archive), and 7 additional two dimensional variables. The MMLEAv2 and a suite of observational data sets have also been regridded onto a 2.5 degree common horizontal grid to reduce data size, and to allow for straightforward model-to-model, and model-to-observations comparison. The MMLEAv2 allows for straightforward comparison between multiple models, and observations, with scope for scientists to download additional data on the native grids from the original data sources to supplement analysis.

The MMLEAv2 is an ensemble of opportunity, including both CMIP5 and CMIP6 class data, with no consistent forcing scenario available for all models. While this provides a limitation for the dataset, possibilities exist to compare warming levels rather than comparing across time (Seneviratne et al., 2021). This enables a more direct comparison between different scenarios, and circumvents this limitation, although it can only be implemented where the warming level itself, not the warming trajectory is important (Hausfather et al., 2022).

An effective way to explore the characteristics of internal variability and forced response in the MMLEAv2 archive is to use the National Science Foundation National Center for Atmospheric Research Climate Variability Diagnostics Package Version6 (CVDPv6; Phillips et al., 2020). Previous versions of this package have been used to investigate how modes of variability are represented over generations of CMIP climate models (Fasullo et al., 2020), questions of model evaluation robustness given the limited duration and certainty in observational datasets (Fasullo et al., 2024), and strengths and weaknesses of US climate models in simulating a broad range of modes of variability (Orbe et al., 2020). In this study, we present the latest version of this package and highlight its value specifically for use with LEs, although it can also be applied to control simulations and models with one realisation. This automated package allows for model comparison with multiple observational data sets, as well as inter-model comparisons, with diagnostics completed over time periods of the user's choosing. The CVDPv6 provides for an ensemble mean view in addition to individual member analysis. The ensemble mean view includes diagnostics of the forced component of climate variability and change, as well as metrics of the rank of the observations within the model ensemble spread to illustrate model bias. The package computes the leading modes of variability in the atmosphere and coupled ocean-atmosphere system including the El Niño Southern Oscillation (ENSO), as well as long-term trends, climatologies, and a variety of climate indices. All results are saved in both graphical and numerical form, allowing for subsequent analysis and display. The user can specify any set of model LEs and observational data sets (including multiple data sets for a given variable) over multiple time periods for analysis. Additionally, the CVDPv6 offers a range of detrending methods that can be applied before computing diagnostics. The versatility of the CVDPv6 makes it an efficient and powerful tool for analysing LE output. In addition, the CVDPv6 displays information from all model simulations and observational data sets on a single page, facilitating model inter-comparisons and assessment of observational uncertainty. In this paper, we will demonstrate the utility of the CVDPv6 as applied to the new MMLEAv2 and highlight some examples of its use and the insights derived therefrom. We note that multiple packages exist that provide ways to investigate climate model output. Examples include: Earth System Model Evaluation Tool (ESMValTool; Eyring et al., 2016), PCMDI Metrics Package (PMP; Lee et al., 2024), and Climate Model Assessment Tool (CMAT; Fasullo, 2020). An earlier version of the CVDP can be run through ESMValTool and the CVDP can also be run from CESM's AMWG Diagnostics Framework (ADF) package and is planned to be run from within the CESM Unified Postprocessing and Diagnostics (CUPiD) package. Amongst many other uses, the CVDP is used alongside the CMAT package to evaluate CESM within the model's development cycle.

The aims of this paper are threefold:

1. Introduce the MMLEAv2 and describe the data available in this extended archive;

2. Demonstrate the utility of the CVDPv6 and the insights one can derive from it;

3. Highlight the importance of using LEs for model evaluation.

2 The MMLEAv2 Dataset

The MMLEAv2 includes 18 models and 16 monthly variables as outlined in Table 1. The original MMLEA consisted of 7 CMIP5-class models with consistent historical and RCP8.5 forcing and 2 models with additional forcing scenarios. The MMLEAv2 contains 6 of the same models from the MMLEA as well as 12 additional CMIP6-class models. While a consistent RCP8.5 future forcing scenario was available for the CMIP5-class models, the CMIP6-class models use one or both of the historical plus SSP370 or SSP585 scenarios. Given that this is an ensemble of opportunity, whichever scenario was available and had the largest number of members is included in the MMLEAv2 (see Table 1). CMIP6-class model data is available for the period 1850-2100, except for GFDL-SPEAR-MED which covers the time period 1921-2100. CMIP5-class data is available for a range of time periods, with the longest simulation from 1850-2099 (MPI-GE) and the shortest from 1950-2100 (CanESM2). The LEs have a minimum of 14 ensemble members for the historical period and a maximum of 100 members. In addition to the original variables in the MMLEA [precipitation (pr), sea level pressure (psl), surface air temperature (tas), horizontal wind stress in the u direction (tauu), horizontal wind stress in the v direction (tauv), sea surface height (zos), 500mb geopotential height (zg500)] the MMLEAv2 provides sea surface salinity (sos), evaporation (evspsbl), mixed layer depth (mlotst), sea-ice concentration (siconc/sic), 20-degree isotherm depth (a proxy for thermocline depth; z20), and 3 monthly extreme indices computed from daily data. We note that unlike the MMLEA, the MMLEAv2 does not provide surface radiative variables. The 3 monthly extreme indices (Zhang et al., 2011) are the monthly maximum of the daily maximum temperature (TXx), the monthly minimum of the daily minimum temperature (TNn), and the monthly maximum of the daily precipitation (Rx1day); they are computed from daily tasmax, tasmin, and pr data using the monmax and monmin functions in Climate Data Operators (CDO; Schulzweida, 2023).

The MMLEAv2 data was compiled from the sources listed below and remapped to a common 2.5 x 2.5 degree grid. CDO conservative mapping was used for pr, psl, tas, tauu, tauv, TXx, TNn, Rx1day and zg500, whereas CDO's distance-weighted mapping was used for all other variables and for all variables in the CESM2 model due to an inability to use the other gridding tools on CESM2's native grid and PSL in GFDL-CM3 due to an additional grididng issue. We note that for CESM2 z20 we regridded the full-depth temperature before calculating the 20-degree isotherm due to a gridding issue, for all other models the 20-degree isotherm was calculated prior to regridding. We also provide remapped observational datasets on the same common grid for ease of use as outlined in Table 2. We note that the data in the CESM models was shifted a month earlier to resolve the issue of netcdf readers reading the data 1 month off, which is an issue on the CESM native temporal grid. If possible, other variables that could be added to the archive in the future are winds at multiple pressure levels in the atmosphere, and surface and top-of-atmosphere fluxes. We note that while not in Table 1 GISS-E2-1-G is also available in the archive for the following variables (ensemble members in brackets for hist/ssp370/ssp585): evspsbl, sos, zos (46/12/15); pr, psl, tas (46/22/10); tauu, tauv (46/17/10); mlotst (46/12/14); tos (40/21/10). We note that this ensemble has multiple physics versions as described by the *p* flag in the ensemble number. This means that this ensemble is a combination of an initial condition and a perturbed parameter ensemble. It has been added to the archive as we believe it's comparison will be useful compared to other ensembles, but note that it should not be treated the same way as the other SMILEs and as such is not included in Table 1 (see https://data.giss.nasa.gov/modelE/cmip6/ for additional details (Kelley et al., 2020).

Data is available at https://rda.ucar.edu/datasets/d651039/ with details of the data found on the project website https://www.cesm.ucar.edu/community-projects/mmlea/v2. The sources for downloading additional variables, or the original data on the native grids are as follows:

- Earth System Grid Federation (ESGF) Nodes for the CMIP6 Archive (https://esgf.github.io/nodes.html); ACCESS-ESM1-5 (Ziehn et al., 2020, 2019), CanESM5 (Swart et al., 2019b, a), EC-Earth3 (Döscher et al., 2022; Wyser et al., 2021; , EC-Earth), IPSL-CM6A-LR (Boucher et al., 2020, 2018), MIROC6 (Tatebe et al., 2019; Tatebe and Watanabe, 2018), MIROC-ES2L (Hajima et al., 2020, 2019; Tachiiri et al., 2019), UKESM1-0-LL (Sellar et al., 2019; Mulcahy et al., 2022) and GISS-E2-1-G (Kelley et al., 2020)

- The MMLEA (https://www.cesm.ucar.edu/community-projects/mmlea); CanESM2 (Kirchmeier-Young et al., 2017), CESM1 (Kay et al., 2015; Deser and Kay, 2014), CSIRO-Mk3-6-0 (Jeffrey et al., 2012), GFDL-CM3 (Sun et al., 2018), MPI-GE (Maher et al., 2019) with additional data available for the following models:
  - CanESM2; https://crd-data-donnees-rdc.ec.gc.ca/CCCMA/products/CanSISE/output/CCCma/CanESM2/
  - CESM1; https://rda.ucar.edu/datasets/d651027/
  - MPI-GE; https://esgf-metagrid.cloud.dkrz.de/; then select project MPI-GE from the top left panel

- The CESM2 Large Ensemble (https://www.cesm.ucar.edu/community-projects/lens2); CESM2 (Rodgers et al., 2021; Danabasoglu et al., 2020))

- The DKRZ node of ESGF (https://esgf-metagrid.cloud.dkrz.de/); MPI-GE-CMIP6 (Olonscheck et al., 2023)

- GFDL SPEAR Large Ensembles (https://www.gfdl.noaa.gov/spear_large_ensembles/); GFDL-SPEAR-MED (Delworth et al., 2020)

- Energy Exascale Earth System Model (E3SM) large ensembles (https://aims2.llnl.gov/search/cmip6/?institution_id=UCSB&?experiment_id=historical,ssp370 and https://portal.nersc.gov/archive/home/c/ccsm/www/E3SMv2/FV1/atm/proc/tseries/month_1); E3SMv1(Stevenson et al., 2023; Bader et al., 2019) and E3SMv2 (Fasullo et al., 2023; Bader et al., 2022)

- The MMLEAv2 Archive (this published archive; https://rda.ucar.edu/datasets/d651039/); GFDL-ESM2M (Burger et al., 2022)

We ask that this paper and the appropriate references from Table 1 (for each model used) are cited when using the MMLEAv2 data.

3 The New CVDPv6

3.1 Scope of the package

The original *Climate Variability Diagnostics Package* (CVDP; Phillips et al., 2014) and the related *Climate Variability and Diagnostic Package for Large Ensembles* (CVDP-LE; Phillips et al., 2020) have been merged into a single application that incorporates the functionality of both packages. This merged application (henceforth referred to as the *CVDPv6*) is an automated analysis and graphics tool that facilitates the exploration of modes of climate variability and change in models, including LEs, and observations. The CVDPv6 is written in NCL (The NCAR Command Language), which can be installed on most commonly used operating systems. The CVDPv6 computes and displays the major modes of climate variability as well as long-term trends and climatologies in models and observations based on a variety of fields including sea surface temperature (sst/tos), surface air temperature (tas), sea level pressure (psl), precipitation (pr), sea ice concentration (sic), sea surface height (zos) and the Atlantic Meridional Overturning Circulation (AMOC). As an analysis tool, it can be used to explore a wide range of topics related to unforced and forced climate variability and change. It can also help with formulating hypotheses, serve as a tool for model evaluation, and generally facilitate curiosity-driven scientific inquiry.

When the CVDPv6 is applied to LEs, it computes metrics that are unique to LEs, for example the ensemble-mean (an estimate of the forced response) and the ensemble spread due to internal variability. The CVDPv6 operates on a user-specified set of model simulations, observational datasets, and time periods, and saves the output (png graphical displays and netcdf data files) to a data repository for later access and further analysis. The CVDPv6 also provides the ability to view the output from two perspectives: *Individual Member* and *Ensemble Summary* (details provided below). A novel feature of the CVDPv6

Table 1. Data included in the MMLEAv2 Dataset. The model and it's reference, forcing type, length of simulation, and variables available are listed in the table. Model names in italics are CMIP6 generation models and forcing, while CMIP5 generation models and forcing are not italicised. The number of ensemble members available for each variable are listed in the table. The variables, scenarios, and set number (sets of LEs we have artificially split the models into for intercomparison purposes) and ensemble size highlighted in italics are those used in the CVDPv6 Tables and Figures. Where an additional variable is listed in brackets the original variable was unavailable and this variable is included in the archive to replace it.

model *CVDP*	**forcing** *CVDPv6* *set no.* *(members)*	**length**	***sst/tos*** sea surface temperature	**zos** sea surface height	***pr*** precip- itation	***psl*** sea level pressure	***tas*** surface air temperature	**tauu** zonal wind stress	**tauv** meridional wind stress	**zg500** 500mb geopotential height	**sos** sea surface salinity	**evspsbl** evapo- oration	**mlotst** mixed layer depth	**TXx** monthly max dailymaxtas	**TNn** monthly min dailymintas	**Rx1day** monthly max pr	**siconc/sic** sea ice conc.	**z20** 20 degree isotherm
1. ACCESS- *ESM1-5* Ziehn et al. (2020)	hist SSP370/585 *Set 1 (40)*	1850- 2100	40	40	40	40 40/10	40	-	-	40	40	40	40	40	40	40	40	40
2. CanESM2 Kirchmeier-Young et al. (2017)	hist RCP8.5 *Set 1 (50)*	1950- 2100	50	-	50	50	50	50	50	50	50	50	50	50	50	50	50	29
3. CanESM5 Swart et al. (2019b)	hist SSP370/585 *Set 1 (25)*	1850- 2100	40 25/25	65 50/50	65 50/50	65 50/50	65 50/50	47 50/50	- -	50 50/50	- -	40 25/25	- -	40 25/25	40 25/25	40 25/25	40 25/25	25 10/9
4. CESM1 Kay et al. (2015)	hist RCP8.5 *Set 1 (40)*	1920- 2100	40	40	40	40	40	40	40	40	40	-	-	40	40	40	40	40
5. CESM2 Rodgers et al. (2021)	hist SSP370 *Set 1 (100)*	1850- 2100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
6. CSIRO- Mk3-6-0 Jeffrey et al. (2012)	hist RCP8.5 *Set 1 (30)*	1850- 2100	30	-	30	30	30	30	30	30	30	30	-	-	-	30	30	30
7. E3SMv1 Stevenson et al. (2023)	hist SSP370 *Set 2 (14)*	1850- 2100	14	13	14	14	14	14	14	14	13	14	14	14	14	-	18	15
8. E3SMv2 Fasullo et al. (2023)	hist SSP370 *Set 2 (21)*	1850- 2100	21 (TS)	21	21	21	21	21	21	21	21	21	-	20	20	20	20	-
9. EC-Earth3 Döscher et al. (2022) Wyser et al. (2021)	hist SSP585 *Set 3 (8)*	1850- 2100	22 58	-	22 58	22 58	23 58	23 8	23 8	22 58	-	24 58	-	18 58	18 58	18 58	22 58	17 4
10. GFDL-CM3 Sun et al. (2018)	hist RCP8.5 *Set 2 (20)*	1920- 2100	20(TS)	-	20	20	20(TS)	20	-	20	20	-	19	-	-	20	20	20
11. GFDL- ESM2M Burger et al. (2022)	hist RCP8.5 *Set 2 (30)*	1861 2100	30	30	30	30	30	30(U10)	30(V10)	-	30	30	30	30	30	-	30	-
12. GFDL- *SPEAR-MED* Delworth et al. (2020)	hist SSP585 *Set 2 (30)*	1921- 2100	30	-	30	30	30	-	-	30	-	-	-	30	30	30	-	-
13. IPSL- *CM6A-LR* Boucher et al. (2020)	hist SSP370/585 *Set 2 (6)*	1850 2100	32 11/6	33 11/7	33 11/7	32 11/6	33 11/7	33 11/7	33 11/7	33 11/6	33 0/8	33 0/8	33 0/8	33 11/6	33 11/6	33 11/6	33 11/6	33 10/-
14. MIROC6 Tatebe et al. (2019)	hist SSP585 *Set 3 (50)*	1850- 2100	50	50	50	50	50	50	50	50	50	50	3	50	50	50	50	46
15. MIROC- *ES2L* Hajima et al. (2020)	hist SSP370/585 *Set 3 (10)*	1850- 2100	30 10/10	31 10/10	31 10/11	31 10/10	31 10/10	31 10/10	31 10/10	30 0/10	31 0/11	31 0/11	- -	30 10/10	30 10/10	30 10/10	30 10/10	46 -/1
16. MPI-GE Maher et al. (2019)	hist RCP8.5 *Set 3 (100)*	1850- 2099	100	100	100	100	100	100	100	100	100	100	-	-	-	-	100	100
17. MPI-GE *CMIP6* Olonscheck et al. (2023)	hist SSP370/585 *Set 3 (30)*	1850 2100	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	50 30	44 30	50 50
18. UKESM1-0-LL Sellar et al. (2019)	hist SSP370/585 *Set 3 (13)*	1850 2100	17 16/5	17 16/5	19 16/5	19 16/5	19 16/5	19 16/5	19 16/5	19 16/5	17 16/5	19 16/5	- -	16 16/5	16 16/5	16 16/5	17 16/5	8 5/4

Table 2. List of observational datasets used within the MMLEAv2 CVDPv6 comparisons. Additional datasets not used in the CVDPv6 are included in italics. Note all tas reference datasets are blended products that use surface air temperature over land and sea surface temperatures over the oceans.

Observational Dataset name	Variables	References
ERSSTv5	tos	Huang et al. (2017)
HadISST v1	tos	Rayner et al. (2003)
ERA5	psl, pr, *Rx1day*, *TXx*, *TNn*	Hersbach et al. (2020)
ERA20C	psl	Poli et al. (2016)
BEST	tas	Rohde and Hausfather (2020)
GISTEMP	tas	Lenssen et al. (2019)
HadCRUT5	tas	Morice et al. (2021)
GPCC	pr	Becker et al. (2013)
GPCP	pr	Adler et al. (2018)
NOAA/NSIDC CDR	siconc	Meier (2021)
ORAS4	zos	Balmaseda et al. (2013)
ORAS5	z20	Copernicus (2021)

is the option to detrend the data using one of 4 methodologies: linear, quadratic, 30-year high pass filter, and for model LEs, the removal of the ensemble mean. While removing the ensemble mean effectively removes the response to the modelled external forcing, the other methodologies are only effective at removing the long-term trend. The inclusion of multiple methods allows for the package to be used as a test-bed for the efficacy of the various observational detrending methods in separating the forced response from internal variability, providing important information about how to interpret the results when the methods are applied to observations. Another novel feature of the CVDPv6 is that the reference data set against which the model LEs are compared can be an observational product or the ensemble mean of an individual LE or multiple LEs at the users discretion. These novel features of the CVDPv6 could be combined, for example, to show the detrended standard deviations of a variable such as surface temperature in future projections with reference to those in the historical period, enabling an assessment of forced changes in the characteristics of the modelled internal variability. To compare the MMLEAv2 archive to the observed historical climate, multiple observational datasets were downloaded and used as reference data within the CVDPv6 comparisons. The observational reference datasets are listed in Table 2. We note that while this data has been regridded as part of the MMLEAv2, the CVDPv6 does not need input files to be on a common grid for its use as the package does regridding on the fly. Compared to the previous version of the CVDP, the new CVDPv6 has some important improvements. It now includes a new Interdecadal Pacific Veriability (IPV) Index (Henley et al., 2015), the Pacific Decadal Variability (PDV) and Atlantic Multidecadal Variability (AMV) definitions have been updated so that the global-mean SST anomaly is no longer removed following Mantua et al. (1997) and Deser and Phillips (2023) respectively. Additionally regression maps for multiple variables have been added for AMV, PDV and both IPV versions and sea surface height (zos) has been added to the list of variables for the climatological averages, standard deviation and trend maps. Finally, a number of new regional time series have been added.

3.2 CVDPv6 output

A list of the output from the CVDPv6 can be found in Table 3. A detailed description of each output field is provided in the Supplemental Materials. All calculations performed in the CVDPv6, including definitions of the modes of variability, are given the *Methodology and Definitions* link at the top of the CVDPv6 output webpage (https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/). A number of summary metrics accompany the graphical displays (see Table 3 for a listing). In particular, for each spatial map, the pattern correlation between the model simulation (which could be an individual member of a LE in the *Individual Member* view or the ensemble-mean of a LE in the *Ensemble Summary* view) and the reference data set (typically observations) is given in the upper right of the map. In addition, for modes of variability based on Empirical Orthogonal Function (EOF) analysis, the fractional variance explained is also provided. The rank maps in the *Ensemble Summary* view, which show the rank of the reference data set within the ensemble spread of the LE, also provide a summary metric of the fractional area of the globe with values between 10% and 90%. For each timeseries displayed in the *Ensemble Summary* view, the ensemble mean and the 25th-75th and 10th-90th percentile ranges across the ensemble members are shown, along with the 10th, 50th and 90th percentile values for linear trends. The percentage of time that the reference dataset lies within the 10th-90th percentile range of the LE values is also shown. The package also produces a synthesis of model performance based on pattern correlations and RMS errors for 15 key spatial metrics of climate variability (listed in Table 3) as well as an overall benchmark based on a mean score of all 15 metrics combined. These summaries are provided in both graphical and tabular format. A novel feature of the Summary Table is the ability to sort the values according to a particular metric (for example, the *North Atlantic Oscillation; NAO*) or by the Mean Score across the 15 metrics. The 10th and 90th percentile values for each metric and Mean Score are also provided in the Tables for each model LE, facilitating model performance intercomparisons. The package also produces a range of output listed in Table 3 that can be viewed in *Individual Member* or in *Ensemble Summary* view. The *Individual Member* view gives the option for each ensemble member to be plotted in its native state or as a bias plot compared to a reference dataset (e.g. observations). The *Ensemble Summary* view plots the ensemble mean, the reference dataset, the difference between the two, and the rank of the reference within the ensemble. If the model and the reference agree well one would expect the reference to sit at a rank inside the model spread i.e. within the 10% and 90% range 80% of the time (Suarez-Gutierrez et al., 2021; Phillips et al., 2020). The rank plot enables a quick evaluation of the variable or metric plotted.

3.3 CVDPv6 applied to MMLEAv2

We have run several applications of the CVDPv6 on the MMLEAv2; output can be found at: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/. In particular, we provide a version without detrending and one where the observations have been quadratically detrended and the LEs have been detrended by removing the ensemble mean as this explicitly removes the forced

Table 3. Contents of the CVDPv6. In addition to the information summarised in this Table, the CVDPv6 displays pattern correlations against the reference dataset and the areal fraction of rank values within 10-90% for all diagnostics plotted in map form. It also provides the temporal fraction of rank values within 10-90% and an ensemble mean summary plot for all diagnostics plotted in timeseries form. For the modes of variability computed using EOF analysis, a graphical summary of the distribution of percent variance explained values across the entire ensemble appears above the numerical values of percent variance explained (10th, 50th and 90th).

Diagnostic	Variables	Season	Plot type
Summary Metrics	all		ensemble table, ensemble graphics, individual table - all for RMS of correlations
ENSO	tas	DJF	metric table
ENSO	psl	DJF	metric table
El Niño Hovmöller	sst	all	metric table
El Niño Hovmöller	sst	all	metric table
AMV Low-Pass	sst	all	metric table
PDV	sst	all	metric table
NAO	psl	JFM	metric table
PNA	psl	DJF	metric table
SAM	psl	DJF	metric table
SST std dev	sst	annual	metric table
PR std dev	pr	annual	metric table
Climatological Averages	sst, tas, psl, pr, sic NH, sic SH, zos	seasonal & annual	maps
Standard Deviations	sst, tas, psl, pr, sic NH, sic SH, zos	seasonal & annual	maps
Linear Trends	SST, sst, tas, psl, pr, sic NH, sic SH, zos	seasonal & annual	maps
Coupled Modes of Variability			
Spatial ENSO	sst, tas, psl, pr, zos	4 seasons	Composite maps and equatorial Pacific Hovmöller diagrams of El Niño, La Niña and their difference
Nino3.4	sst	monthly	timeseries, standard deviations, spectra, wavelets, autocorrelation, running 30-year standard deviation timeseries
AMV, PDV, IPV	sst, tas, pr	monthly	regression mapsm, timeseries, spectra
AMOC	sst, tas, overturning	monthly	climatological means, standard deviation, patterns, timeseries, spectra, lag correlations
Atmospheric Modes of Variability			
SO, NAM, NAO, SAM, PNA, NPO, PSA, PSA2	psl	seasonal & annual	timeseries, regression maps
Global Timeseries	sst, tas, pr, prland	seasonal & annual	timeseries
Regional Timeseries			
Various	sst, psl	monthly	timeseries
Sea Ice Extent (NH & SH)	sie	seasonal, annual, monthly, climatology	timeseries

response from each LE (note that detrending is omitted from the calculations of climatologies and trends). The CVDPv6 comparisons are divided into 3 sets of models (see Table 1 for which model is in which set) for ease of visualisation. The CVDPv6 is run on each of the 3 sets of models for the time periods 1950-2022 and 2027-2099. An additional set named Set123 only includes the Summary Metrics (tables and graphics) to facilitate an intermodel comparison across the entirety of the MMLEAv2. All analysis is completed on RCP8.5 and SSP5-8.5 except for GISS-E2-G, CESM2 and UKESM1-0-LL where SSP3-7.0 is used. Observational datasets used for these comparisons are shown in Table 2.

4 Intermodel Comparison Tables & Graphics

This section demonstrates the intermodel comparison tables and graphics from the CVDPv6 for all MMLEAv2 ensemble members and types of conclusions that can be drawn from them.

Multiple evaluation methods exist to compare climate variability in models with observations. While they do not always provide consistent answers, together they can be used to create a more general picture of model performance. Here, we use two complementary evaluation methods; pattern correlations (Figure 1) and root mean square differences (RMS; Figure 2). The pattern correlation informs about the spatial similarity between two maps, while the RMS provides a summary of the relative amplitude of the maps. For the pattern correlation mean score, CESM2 and GFDL-SPEAR have the highest individual ensemble member scores compared to observations (around 0.9), followed by a group of models clustered with their highest correlation in the 0.8-0.85 range. For RMS, the highest performance (lowest RMS values) are found again for GFDL-SPEAR, but also EC-Earth3 and UKESM1-0-LL. These results demonstrate that GFDL-SPEAR is a high performer for both spatial pattern and amplitude, however, it is worth keeping in mind that the evaluation is also metric-dependent. For example, ACCESS-ESM1-5 has one of the lowest RMS scores and high correlations for ENSO TAS in DJF showing it performs well in both pattern and amplitude for this metric. However, for the La Niña Hovmöller, it is in the bottom two models for the spatial pattern. Another example is GFDL-CM3, which has the lowest pr STD correlation, but conversely performs relatively well on amplitude for the PNA metric. This demonstrates that models can perform well for one metric and evaluation type, but poorly for another, meaning that one must be careful in selecting an evaluation method and metric fit for purpose for any given study.

There are some metrics where all models tend to perform well or poorly. For example, the red columns of Figure 2, PDV, SST STD and PR STD, all show low error compared to observations in all models and ensemble members for RMS. Conversely, the blue-coloured columns such as ENSO PSL and AMV Low Pass show poor performance across all ensemble members of all models. For the pattern correlation, again all models and members perform poorly for the AMV Low Pass where there is a maximum correlation of 0.6 (Figure 1). On the contrary, no ensemble members of any model has a correlation below 0.88 for the SAM, highlighting high performance of all models for this metric. Metrics where all models perform poorly can tell us about general biases found in all climate models. The AMV Low Pass is an example of where the models perform poorly in both the spatial pattern and amplitude, athough this could be due to the tendency of the AMV Low Pass index to mix together multiple independent processes (Wills et al., 2019; Deser and Phillips, 2023; O'Reilly et al., 2023) which might have different relative weights in models and observations. On the flip side the low RMS for SST STD highlights that the amplitude of SST variability is generally correct in all models, while in many models the maximum correlation is below 0.8 and as low as 0.4 suggesting that while the amplitude is correct there is bias in the spatial pattern. Teasing out where consistent model biases exist and whether they appear for both pattern and amplitude can inform scientific research and model development to improve models in the coming model generations.

Ensemble members from the same model with the same external forcing can have a wide range of performance compared with observations. For metrics such as the AMV Low Pass, the pattern correlation with observations ranges from slightly negative to 0.6 in an individual model. ENSO TAS also has a range of possible correlations that vary by up to 0.3 between members These examples highlight that for metrics such as these two, the use of a single ensemble member would not give a correct model evaluation. For other metrics such as PR STD, the range of correlations within a single model only varies by a magnitude of 0.1, demonstrating that in this case a large ensemble is not necessary for model comparison to observations. The range of correlations across ensemble members is linked to the magnitude of internal variability of a metric relative to its ensemble-mean (e.g. Milinski et al., 2020; Lee et al., 2021). This range is not determined solely by ensemble size as demonstrated by the SAM metric where the 100-member CESM2 model has a much larger range of possible correlations than the 100-member MPI-GE. This difference in range from two models with the same ensemble size does, however, demonstrate that the variability of a single metric can be model-dependent similar to Lehner et al. (2020). Another example that highlights the importance of considering the full ensemble, rather than one member is the El Niño Hovmöller, which for CESM2 can have a RMS as small as 0.49 and as large as 0.76. If the member with the 0.76 RMS was used this model would be deemed poor at representing the El Niño evolution, while the opposite conclusion would be made if the 0.49 member was selected. Similar results are again found for smaller ensembles with the PNA in ACCESS-ESM1-5 as low as 0.43 or as high as 0.68 depending on which member was chosen. In general, Figures 1 and 2 highlight the importance of using LEs for model evaluation, particularly for metrics with high internal climate variability.

5 Model Evaluation Figures

This section gives examples of Figures output from the CVDPv6 and associated insights that can be made into the MMLEAv2 models.

The CVDPv6 can be used to assess annual global average surface temperature and determine whether the response to external forcing (i.e. greenhouse gases, anthropogenic aerosols and volcanic forcing) in combination with internal variability is similar to observations. For the Set 1 MMLEAv2 models shown in Figure 3, only CESM2 and CESM1 largely encompass the observations within the model spread as highlighted by the temporal ranks (e.g. the percentage of time that the observed value lies within the 10th-90th percentile range of the model's ensemble members shown in the bottom left of the plots), where these models have 85% and 79% respectively. ACCESS-ESM1-5, CanESM2 and CanESM5 all have a larger climate sensitivity than is estimated from observations, as demonstrated by all members warming much more than the observations at the end of the historical period. The CSIRO-Mk360 model has a bias in the historical period where it does not warm from 1960-2000, dissimilar to what is observed. In response to volcanic forcing, all models have a distinctive dip in temperature in 1963 (Agung), 1982 (El Chichón), and 1991 (Pinautubo) that is similar in magnitude to the observed record, except CSIRO-Mk36 which does dip in response to eruptions but at a different rate to observations. This suggests that the global-mean annual-mean temperature response to large tropical volcanic eruptions is realistic in most Set 1 models.

Accurately representing large-scale modes of climate variability in models is important as these modes are key components of the climate system. The large-scale modes shown in the following examples are calculated after the removal of the ensemble mean. Hence they approximate pure internal variability with all external trends removed. This is a key use of LEs and a powerful tool within the CVDPv6. We note that changes in the variability itself due to external forcing are not removed using this methodology, so changes in the variability itself can be assessed using this method. The CVDPv6 uses a rank histogram approach where the rank of observations is shown within each model's spread. This type of comparison between observations and models is only possible using LEs. The NAO pattern in DJF for the Set 1 models is shown in Figure 4. CESM2 and CanESM5 have the largest percentage of white area (indicating that observations lie in the center of the distribution of the model's diagnosed internal variability patterns) in the rank histogram map (areal percentage of observed values lying within the 10th-90th percentile of the LE values of 61% and 59% respectively), while CSIRO-Mk36 has the lowest at 28%. For the most part, the Set 1 models tend to have similar biases with the NAO variability overestimated in the North Pacific and over Alaska, and underestimated in the North Atlantic and over most of Russia and Siberia. A similar plot for the PDV is shown in Figure 5 for Set 3 models. In this case, the pattern biases are quite different for different models, especially in the tropical Pacific, where some models have too much PDV variability and some have too little. There is a consistent bias in the North Pacific, where all Set 3 models show too large Kuroshio-Oyashio Extension anomalies that are located too far north. EC-Earth3, MPI-GE and MPI-GE-CMIP6 are the models with more than 60% of the areal percentable of the ranks occurring between the 10th and 90th percentile, indicating that they capture observed PDV best out of the Set 3 models. Modes of variability can also be assessed by their spectra (Figure 6; ENSO spectra). For the ENSO spectra in Set 2 models, we find that GFDL-CM3, E3SMv2 and IPSL-CM6A have a peak that sits at 3 years, while observations have a broader 3-7 year peak. GFDL-ESM2M is similar to observations but with a stronger peak as is GISS-E2-G with an even stronger peak. GFDL-SPEAR and E3SMv1 however, are much more in agreement with observations, with the ensemble spread encompassing observations.

Another important aspect of model evaluation is the consideration of how well a model represents the observed impacts of a mode of variability. In Figure 7 we demonstrate the SON precipitation impacts from La Niña in the Set 3 models. MIROC6 has 70% of the areal percentable of ranks between the 10th and 90th percentile, with all other models above 62%. This shows that over 60% of the globe has rainfall impacts in SON that are well represented in the Set 3 models. Be that as it may, biases do exist. For example, all models have a wet bias in the western tropical Pacific, and the Alaskan coastline and a dry bias over California in relation to observed La Niña events. In other regions models have differing biases with the La Niña teleconnection to Australia overestimated in some models and underestimated in others. Whether these biases are due to model errors or uncertainty in the observed composite due to limited sampling of events (Deser et al., 2018, 2017) and/or data issues is less clear due to a sparsity of observations in some regions. While not shown in this paper, the CVDPv6 allows for comparison with multiple observational products, which can help characterize observational uncertainty.

While we cannot directly compare future projections with observations, we can compare models with each other and with the multi-model or in this case multi-ensemble mean. The CVDPv6 outputs data in netcdf files, which can then be used to create new plots. We created Figure 8 from these outputs. It shows (for Set 2 models) an individual model comparison to the multi-ensemble mean (MEM; calculated across all models from the 3 sets). This method of model intercomparison highlights where the forced response differs across models. We find that the trend in temperature in GFDL-ESM2M is smaller than the MEM, IPSL-CM6A and E3SMv1 have a larger trend than the MEM, and GFDL-CM3 has a larger trend everywhere but Antarctica and the Southern Ocean where it is smaller. E3SMv2 and GFDL-SPEAR have similar trends to the MEM. In this case, differences from the MEM largely show differences in global warming between models, due to differences in climate sensitivity and emissions scenario, but a similar comparison for other variables would help to characterize other aspects of the structural uncertainty in climate projections. Interestingly there are substantial regional changes in the rank of each model within the ensemble spread. While both E3SMv1 and v2 sit at high ranks in the higher latitudes, and GFDL-ESM2M sits at low ranks for all regions, IPSL-CM6A sits at a high rank in parts of the extratropics and low in the tropics while GFDL-CM3 has low ranks only in the Southern Ocean and GFDL-SPEAR has low ranks in specific locations in the extratropical northern hemisphere. This means that regional warming is not solely dependent on the magnitude of global warming in these models. We note that this type of comparison could also be done at warming levels which would present a fairer comparison across the varying future scenarios. This is, however, not possible with the CVDPv6 package alone. While such analysis allows comparison between models it does not enable us to assess which model is most realistic. Research using emergent constraints from observations could be applied to the MMLEAv2 to answer such a question.

To highlight the use of the extreme indices available in the MMLEAv2 dataset we show the change at the end of the century (2090-2099) compared to the historical period (1950-1959) of the monthly maximum of daily maximum temperature (TXx) in June, July, August (Figure 9). For a fairer comparison of the varying future scenarios (compared to Figure 8) we only use 8 models that have the SSP370 future scenario available for this variable. In the SSP370 future scenario, all models show an increase in TXx over land (and over almost all of the ocean). The magnitude of this change is, however, model dependent. E3SMv1 shows the largest increase followed by UKESM1-0-LL, with MPI-GE-CMIP6 and MIROC-E3SL showing the lowest increases. The range of increases in TXx over land is greater than 10°C, highlighting how variable the future projections of this extreme metric are across LEs. Figure 3b in (Deser et al., 2020) reported a similar finding for daily July heat extremes at the grid box containing Dallas Texas.

6 Conclusions

This work has presented two new complementary resources for the study of climate variability and change: the MMLEAv2 and the CVDPv6. Designed for ease-of-use, these tools provide a broad synthesis of internal and forced contributions to the major modes of climate variability and trends in model LEs in relation to observations, facilitating model evaluation, and intercomparison. Here, we have demonstrated some of the insights that can be derived by applying the CVDPv6 to the MMLEAv2. In particular, we have highlighted the following points:

- the MMLEAv2 is an extension of the MMLEA (Deser et al., 2020) with additional models and variables available (18 models and 16 two-dimensional variables)
- the MMLEAv2 is remapped onto a common 2.5 x 2.5 degree grid for ease of inter-model comparison
- The MMLEAv2 also provides observed reference datasets on the common 2.5 x 2.5 degree grid for ease of model-to-observation comparison
- The CVDPv6 provides a powerful and efficient way to analyse LE output, including the MMLEAv2
- A preliminary model evaluation has been completed using the CVDPv6 applied to MMLEAv2 that highlights common model biases, and which models perform well or poorly compared to observations
- Figures 1 & 2 of this paper allow the user to determine which climate variables need a large ensemble for fair comparison against observations
- The CVDPv6 enables the exploration of climate model output and observations, including modes of variability, time-series, trends, and spatial maps of key variables

- Multiple detrending methods can be applied in the CVDPv6, an improvement on previous versions of the package allowing for the efficacy of methods such as quadratic detrending that are typically applied to single realisations (such as observations) to be tested against the removal of the ensemble mean approach (effectively removing effects of external forcing)

- The CVDPv6 can use any reference dataset (model or observations) and time period for comparison

- the CVDPv6 output is provided in two complementary ways: *Individual Member* view and in *Ensemble Summary* view

- *Ensemble Summary* view includes not just spatial plots and timeseries, but also a rank plot of where the observations sit within the model spread for easy evaluation (i.e. white areas show high model-to-observational agreement)

- Netcdf output is also available from the CVDPv6, which allows uses to create their own additional figures from diagnostics computed in the package.

- Graphical output is available from the CVDPv6 which can be output as png files for publications and presentation

- Output is available as an easily navigable HTML format for users to click through (i.e. https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/)

Additionally, we have demonstrated the utility of LEs for model evaluation. Observations can be considered as one realisation of the world that we live in, meaning that observations are best compared with multiple realisations from a climate model, such as a LE. This is highlighted by the spread of evaluation metrics found across the ensemble. In some cases one ensemble member would evaluate poorly against observations while another would compare favourably. For this reason LEs are important for model evaluation, especially for highly variable quantities. Overall, the MMLEAv2 will allow for exciting new science using LEs and the CVDPv6 is a powerful tool made specifically for analysis of LEs and their unique characteristics.

Code and data availability. The CVDPv6 code is available at: https://www.cesm.ucar.edu/projects/cvdp/code The MMLEAv2 data is available at https://rda.ucar.edu/datasets/d651039/ with details of the data found on the project website https://www.cesm.ucar.edu/community-projects/mmlea/v2. All processed data, the script used to calculate z20, and a frozen version of the CVDPv6 code used in this publication are available on zenodo at 10.5281/zenodo.15589035 https://zenodo.org/records/15589035

Author contributions. NM led the compilation of the MMLEAv2 and wrote the manuscript. AP led the CVDPv6 Package Development and ran the package on the MMLEAv2 data. CD was key in the intellectual development of the CVDPv6 and contributed to the writing of Section 3. NM, AP & CD designed the manuscript. RW was key in adding data to the MMLEAv2 archive and initiating it's existence. FL compiled the MMLEA original archive. JF is a CVDP super-user who has helped develop the CVDPv6 over many years. JC specifically helped with the testing and development of CVDPv6. LB & UB set up the cmip-ng archive at ETH that provided much data for the MMLEA. JJ processed and quality checked the z20 field. All authors contributed to the revision of this manuscript.

Competing interests. The authors declare no competing interests are present.

Acknowledgements. We would like to acknowledge computing support from the Casper system (https://ncar.pub/casper) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. NM and JJ were supported by the Australian Research Council Discovery Early Career Researcher Award DE230100315. NM also acknowledges support from NASA (grant 80NSSC23K0358) and the CIRES Visiting Fellows Program and the NOAA Cooperative Agreement with CIRES, NA17OAR4320101. We thank C2SM for providing/updating cmip6-ng where we obtained data (already regridded to a 2.5 degree grid by Brunner et al., 2020), and Ruth Lorenz for providing the zos fields as part of this archive. We also thank Dillon Amaya for his work in quality-checking the datasets. We additionally thank Thomas Frölicher and Dirk Olonscheck for providing data and access for GFDL-ESM2M and MPI-GE-CMIP6 respectively. The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation under Cooperative Agreement 1852977. The efforts of JF, JMC and FL are supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) under Award Number DE-SC0022070. Dr. Fasullo was also supported by NSF Award 2103843. RCJW was supported by the Swiss National Science Foundation (Award PCEFP2 203376). LB has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2037 "CLICCS – Climate, Climatic Change, and Society" – project no. 390683824, a contribution to the Center for Earth System Research and Sustainability (CEN) of the University of Hamburg. This research/project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We also thank the anonymous reviewers for their time and effort.

References

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D.-B.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, Atmosphere, 9, https://doi.org/10.3390/atmos9040138, 2018.

Bader, D. C., Leung, R., Taylor, M., and McCoy, R. B.: E3SM-Project E3SM1.0 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.2294, 2019.

Bader, D. C., Leung, R., Taylor, M., and McCoy, R. B.: E3SM-Project E3SM2.0 model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.16953, 2022.

Balmaseda, M. A., Mogensen, K., and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, Quarterly Journal of the Royal Meteorological Society, 139, 1132–1161, https://doi.org/https://doi.org/10.1002/qj.2063, 2013.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, Earth System Science Data, 5, 71–99, https://doi.org/10.5194/essd-5-71-2013, 2013.

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., Cadule, P., Devilliers, M., Ghattas, J., Lebas, N., Lurton, T., Mellul, L., Musat, I., Mignot, J., and Cheruy, F.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.1534, 2018.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 010, https://doi.org/https://doi.org/10.1029/2019MS002010, e2019MS002010 10.1029/2019MS002010, 2020.

Brunner, L., Hauser, M., Lorenz, R., and Beyerle, U.: The ETH Zurich CMIP6 next generation archive: technical documentation, ETH Zürich, https://iacweb.ethz.ch/staff/lukbrunn/welcome/files/Brunner2020.pdf, 2020.

Burger, F. A., Terhaar, J., and Frölicher, T. L.: Compound marine heatwaves and ocean acidity extremes, Nature Communications, 13, 4722, https://doi.org/10.1038/s41467-022-32120-7, 2022.

Copernicus: ORAS5 global ocean reanalysis monthly data from 1958 to present., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/DOI: 10.24381/cds.67e8eeb7, 2021.

Danabasoglu, G., Deser, C., Rodgers, K., and Timmermann, A.: CESM2 Large Ensemble, https://doi.org/https://doi.org/10.26024/kgmp-c556, 2020.

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., Ginoux, P., Gudgel, R., Hallberg, R. W., Harris, L., Harrison, M. J., Johnson, N., Kapnick, S. B., Lin, S.-J., Lu, F., Malyshev, S., Milly, P. C., Murakami, H., Naik, V., Pascale, S., Paynter, D., Rosati, A., Schwarzkopf, M., Shevliakova, E., Underwood, S., Wittenberg, A. T., Xiang, B., Yang, X., Zeng, F., Zhang, H., Zhang, L., and Zhao, M.: SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 895, https://doi.org/https://doi.org/10.1029/2019MS001895, 2020.

Deser, C. and Kay, J.: CESM1 CAM5 BGC 20C + RCP8.5 large ensemble data, including the lossy data compression project., https://doi.org/https://doi.org/10.5065/d6j101d1, 2014.

Deser, C. and Phillips, A. S.: Spurious Indo-Pacific Connections to Internal Atlantic Multidecadal Variability Introduced by the Global Temperature Residual Method, Geophysical Research Letters, 50, e2022GL100 574, https://doi.org/https://doi.org/10.1029/2022GL100574, e2022GL100574 2022GL100574, 2023.

Deser, C., Simpson, I. R., McKinnon, K. A., and Phillips, A. S.: The Northern Hemisphere Extratropical Atmospheric Circulation Response to ENSO: How Well Do We Know It and How Do We Evaluate Models Accordingly?, Journal of Climate, 30, 5059 – 5082, https://doi.org/10.1175/JCLI-D-16-0844.1, 2017.

Deser, C., Simpson, I. R., Phillips, A. S., and McKinnon, K. A.: How Well Do We Know ENSO's Climate Impacts over North America, and How Do We Evaluate Models Accordingly?, Journal of Climate, 31, 4991 – 5014, https://doi.org/10.1175/JCLI-D-17-0783.1, 2018.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth system model initial-condition large ensembles and future prospects, Nature Climate Change, 10, 277–286, https://doi.org/10.1038/s41558-020-0731-2, 2020.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geoscientific Model Development, 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

(EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.181, 2019.

Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I., Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K.-D., Hagemann, S., Juckes, M., Kindermann, S., Krasting, J., Kunert, D., Levine, R., Loew, A., Mäkelä, J., Martin, G., Mason, E., Phillips, A. S., Read, S., Rio, C., Roehrig, R., Senftleben, D., Sterl, A., van Ulft, L. H., Walton, J., Wang, S., and Williams, K. D.: ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP, Geoscientific Model Development, 9, 1747–1802, https://doi.org/10.5194/gmd-9-1747-2016, 2016.

Fasullo, J., Golaz, J.-C., Caron, J., Rosenbloom, N., Meehl, G., Strand, W., Glanville, S., Stevenson, S., Molina, M., Shields, C., Zhang, C., Benedict, J., and Bartoletti, T.: An Overview of the E3SM version 2 Large Ensemble and Comparison to other E3SM and CESM Large Ensembles, EGUsphere, 2023, 1–32, https://doi.org/10.5194/egusphere-2023-2310, 2023.

Fasullo, J. T.: Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1), Geoscientific Model Development, 13, 3627–3642, https://doi.org/10.5194/gmd-13-3627-2020, 2020.

Fasullo, J. T., Phillips, A. S., and Deser, C.: Evaluation of Leading Modes of Climate Variability in the CMIP Archives, Journal of Climate, 33, 5527 – 5545, https://doi.org/10.1175/JCLI-D-19-1024.1, 2020.

Fasullo, J. T., Caron, J. M., Phillips, A., Li, H., Richter, J. H., Neale, R. B., Rosenbloom, N., Strand, G., Glanville, S., Li, Y., Lehner, F., Meehl, G., Golaz, J.-C., Ullrich, P., Lee, J., and Arblaster, J.: Modes of Variability in E3SM and CESM Large Ensembles, Journal of Climate, 37, 2629 – 2653, https://doi.org/10.1175/JCLI-D-23-0454.1, 2024.

Goldenson, N., Thackeray, C. W., Hall, A. D., Swain, D. L., and Berg, N.: Using Large Ensembles to Identify Regions of Systematic Biases in Moderate-to-Heavy Daily Precipitation, Geophysical Research Letters, 48, e2020GL092 026, https://doi.org/https://doi.org/10.1029/2020GL092026, e2020GL092026 2020GL092026, 2021.

Hajima, T., Abe, M., Arakawa, O., Suzuki, T., Komuro, Y., Ogura, T., Ogochi, K., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Ohgaito, R., Ito, A., Yamazaki, D., Ito, A., Takata, K., Watanabe, S., Kawamiya, M., and Tachiiri, K.: MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical, https://doi.org/10.22033/ESGF/CMIP6.5602, 2019.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geoscientific Model Development, 13, 2197–2244, https://doi.org/10.5194/gmd-13-2197-2020, 2020.

Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., and Zelinka, M.: Climate simulations: recognize the ‘hot model’ problem, Nature, 605, 26–29, https://doi.org/https://doi.org/10.1038/d41586-022-01192-2, 2022.

Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Geophysical Research Letters, 39, https://doi.org/https://doi.org/10.1029/2011GL050087, 2012.

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., and Folland, C. K.: A Tripole Index for the Interdecadal Pacific Oscillation, Climate Dynamics, 45, 3077–3090, https://doi.org/10.1007/s00382-015-2525-1, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons, Journal of Climate, 30, 8179 – 8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, Journal of Marine Systems, 78, 28–41, https://doi.org/https://doi.org/10.1016/j.jmarsys.2008.11.026, 2009.

Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., Wong, K., and Syktus, J.: Australia’s CMIP5 submission using the CSIRO-Mk3.6 model, Australian Meteorological and Oceanographic Journal, 63, 1–13, 2012.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, Bulletin of American Meteorological Society, 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Miller, R. L., Murray, L. T., Oinas, V., Orbe, C., García-Pando, C. P., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., and Yao, M.-S.: GISS-E2.1: Configurations and Climatology, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 025, https://doi.org/https://doi.org/10.1029/2019MS002025, e2019MS002025 10.1029/2019MS002025, 2020.

Kirchmeier-Young, M., Zwiers, F., and Gillett, N.: Attribution of Extreme Events in Arctic Sea Ice Extent, Journal of Climate, 30, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1, 2017.

Labe, Z. M. and Barnes, E. A.: Comparison of Climate Model Large Ensembles With Observations in the Arctic Using Simple Neural Networks, Earth and Space Science, 9, e2022EA002 348, https://doi.org/https://doi.org/10.1029/2022EA002348, e2022EA002348 2022EA002348, 2022.

Lee, J., Planton, Y. Y., Gleckler, P. J., Sperber, K. R., Guilyardi, E., Wittenberg, A. T., McPhaden, M. J., and Pallotta, G.: Robust Evaluation of ENSO in Climate Models: How Many Ensemble Members Are Needed?, Geophysical Research Letters, 48, e2021GL095 041, https://doi.org/https://doi.org/10.1029/2021GL095041, e2021GL095041 2021GL095041, 2021.

Lee, J., Gleckler, P. J., Ahn, M.-S., Ordonez, A., Ullrich, P. A., Sperber, K. R., Taylor, K. E., Planton, Y. Y., Guilyardi, E., Durack, P., Bonfils, C., Zelinka, M. D., Chao, L.-W., Dong, B., Doutriaux, C., Zhang, C., Vo, T., Boutte, J., Wehner, M. F., Pendergrass, A. G., Kim, D., Xue, Z., Wittenberg, A. T., and Krasting, J.: Systematic and objective evaluation of Earth system models: PCMDI Metrics Package (PMP) version 3, Geoscientific Model Development, 17, 3919–3948, https://doi.org/10.5194/gmd-17-3919-2024, 2024.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth System Dynamics, 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP Uncertainty Model, Journal of Geophysical Research: Atmospheres, 124, 6307–6326, https://doi.org/https://doi.org/10.1029/2018JD029522, 2019.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability, Journal of Advances in Modeling Earth Systems, 11, 2050–2069, https://doi.org/10.1029/2019MS001639, 2019.

Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, Earth System Dynamics, 12, 401–418, https://doi.org/10.5194/esd-12-401-2021, 2021a.

Maher, N., Power, S. B., and Marotzke, J.: More accurate quantification of model-to-model agreement in externally forced climatic responses over the coming century, Nature Communications, 12, 788, https://doi.org/10.1038/s41467-020-20635-w, 2021b.

Maher, N., Wills, R. C. J., DiNezio, P., Klavans, J., Milinski, S., Sanchez, S. C., Stevenson, S., Stuecker, M. F., and Wu, X.: The future of the El Niño–Southern Oscillation: using large ensembles to illuminate time-varying responses and inter-model differences, Earth System Dynamics, 14, 413–431, https://doi.org/10.5194/esd-14-413-2023, 2023.

Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C.: A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production*, Bulletin of the American Meteorological Society, 78, 1069 – 1080, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2, 1997.

Meier, W. N., F.-F. W. A. K. . S. J. S.: NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4, https://doi.org/10.7265/efmz-2t65, 2021.

Milinski, S., Maher, N., and Olonscheck, D.: How large does a large ensemble need to be?, Earth System Dynamics, 11, 885–901, https://doi.org/10.5194/esd-11-885-2020, 2020, 2020.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/https://doi.org/10.1029/2019JD032361, e2019JD032361 2019JD032361, 2021.

Mulcahy, J., Rumbold, S., Tang, Y., Walton, J., Hardacre, C., Stringer, M., Hill, R., Kuhlbrodt, T., and Jones, C.: MOHC UKESM1.1-LL model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.16781, 2022.

Olonscheck, D., Suarez-Gutierrez, L., Milinski, S., Beobide-Arsuaga, G., Baehr, J., Fröb, F., Ilyina, T., Kadow, C., Krieger, D., Li, H., Marotzke, J., Plésiat, , Schupfner, M., Wachsmann, F., Wallberg, L., Wieners, K.-H., and Brune, S.: The New Max Planck Institute Grand Ensemble With CMIP6 Forcing and High-Frequency Model Output, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 790, https://doi.org/https://doi.org/10.1029/2023MS003790, e2023MS003790 2023MS003790, 2023.

Orbe, C., Roekel, L. V., Ángel F. Adames, Dezfuli, A., Fasullo, J., Gleckler, P. J., Lee, J., Li, W., Nazarenko, L., Schmidt, G. A., Sperber, K. R., and Zhao, M.: Representation of Modes of Variability in Six U.S. Climate Models, Journal of Climate, 33, 7591 – 7617, https://doi.org/10.1175/JCLI-D-19-0956.1, 2020.

O'Reilly, C. H., Patterson, M., Robson, J., Monerie, P. A., Hodson, D., and Ruprich-Robert, Y.: Challenges with interpreting the impact of Atlantic Multidecadal Variability using SST-restoring experiments, npj Climate and Atmospheric Science, 6, 14, https://doi.org/10.1038/s41612-023-00335-0, 2023.

Phillips, A. S., Deser, C., and Fasullo.: A New Tool for Evaluating Modes of Variability in Climate Models, EOS, 95, 453–455, https://doi.org/doi: 10.1002/2014EO490002, 2014.

Phillips, A. S., Deser, C., Fasullo, J., P., S. D., and Simpson, I. R.: Assessing Climate Variability and Change in Model Large Ensembles: A User's Guide to the "Climate Variability Diagnostics Package for Large Ensembles" Version 1.0., https://doi.org/doi:10.5065/h7c7-f961, 2020.

Planton, Y. Y., Lee, J., Wittenberg, A. T., Gleckler, P. J., Guilyardi, , McGregor, S., and McPhaden, M. J.: Estimating Uncertainty in Simulated ENSO Statistics, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 147, https://doi.org/https://doi.org/10.1029/2023MS004147, e2023MS004147 2023MS004147, 2024.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, Journal of Climate, 29, 4083 – 4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/https://doi.org/10.1029/2002JD002670, 2003.

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J.-E., Simpson, I. R., Stein, K., Stuecker, M. F., Yamaguchi, R., Bódai, T., Chung, E.-S., Huang, L., Kim, W. M., Lamarque, J.-F., Lombardozzi, D. L., Wieder, W. R., and Yeager, S. G.: Ubiquity of human-induced changes in climate variability, Earth System Dynamics, 12, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021, 2021.

Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth System Science Data, 12, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020, 2020.

Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, npj Climate and Atmospheric Science, 1, 28, https://doi.org/10.1038/s41612-018-0038-4, 2018.

Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Ilyina, T., Dunne, J. P., Takano, Y., Christian, J. R., Long, M. C., Frölicher, T. L., Slater, R., and Lehner, F.: Time of Emergence and Large Ensemble Intercomparison for Ocean Biogeochemical Trends, Global Biogeochemical Cycles, 34, e2019GB006 453, https://doi.org/https://doi.org/10.1029/2019GB006453, e2019GB006453 2019GB006453, 2020.

Schulzweida, U.: CDO User Guide (2.3.0), Zenodo, https://doi.org/10.5281/zenodo.10020800, 2023.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, https://doi.org/https://doi.org/10.1029/2019MS001739, 2019.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate Supplementary Material, Availablefromhttps://www.ipcc.ch/, 2021.

Simpson, I. R., Shaw, T. A., Ceppi, P., Clement, A. C., Fischer, E., Grise, K. M., Pendergrass, A. G., Screen, J. A., Wills, R. C. J., Woollings, T., Blackport, R., Kang, J. M., and Po-Chedley, S.: Confronting Earth System Model trends with observations, Science Advances, 11, eadt8035, https://doi.org/10.1126/sciadv.adt8035, 2025.

Stevenson, S., Huang, X., Zhao, Y., Di Lorenzo, E., Newman, M., van Roekel, L., Xu, T., and Capotondi, A.: Ensemble Spread Behavior in Coupled Climate Models: Insights From the Energy Exascale Earth System Model Version 1 Large Ensemble, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 653, https://doi.org/https://doi.org/10.1029/2023MS003653, e2023MS003653 2023MS003653, 2023.

Suarez-Gutierrez, L., Milinski, S., and Maher, N.: Exploiting large ensembles for a better yet simpler climate model evaluation, Climate Dynamics, 57, 2557–2580, https://doi.org/10.1007/s00382-021-05821-w, 2021.

Sun, L., Alexander, M., and Deser, C.: Evolution of the Global Coupled Climate Response to Arctic Sea Ice Loss during 1990-2090 and Its Contribution to Climate Change, Journal of Climate, 31, 7823–7843, 2018.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.1303, 2019a.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geoscientific Model Development, 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019b.

Tachiiri, K., Abe, M., Hajima, T., Arakawa, O., Suzuki, T., Komuro, Y., Ogochi, K., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Ohgaito, R., Ito, A., Yamazaki, D., Ito, A., Takata, K., Watanabe, S., and Kawamiya, M.: MIROC MIROC-ES2L model output prepared for CMIP6 ScenarioMIP, https://doi.org/10.22033/ESGF/CMIP6.936, 2019.

Tatebe, H. and Watanabe, M.: MIROC MIROC6 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.881, 2018.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geoscientific Model Development, 12, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019, 2019.

Weisheimer, A., Baker, L. H., Bröcker, J., Garfinkel, C. I., Hardiman, S. C., Hodson, D. L. R., Palmer, T. N., Robson, J. I., Scaife, A. A., Screen, J. A., Shepherd, T. G., Smith, D. M., and Sutton, R. T.: The Signal-to-Noise Paradox in Climate Forecasts: Revisiting Our Understanding and Identifying Future Priorities, Bulletin of the American Meteorological Society, 105, E651 – E659, https://doi.org/10.1175/BAMS-D-24-0019.1, 2024.

Wills, R. C. J., Armour, K. C., Battisti, D. S., and Hartmann, D. L.: Ocean–Atmosphere Dynamical Coupling Fundamental to the Atlantic Multidecadal Oscillation, Journal of Climate, 32, 251 – 272, https://doi.org/10.1175/JCLI-D-18-0269.1, 2019.

Wills, R. C. J., Deser, C., McKinnon, K. A., Phillips, A., Po-Chedley, S., Sippel, S., Merrifield, A. L., Bône, C., Bonfils, C., Camps-Valls, G., Cropper, S., Connolly, C., Duan, S., Durand, H., Feigin, A., Fernandez, M. A., Gastineau, G., Gavrilov, A., Gordon, E., Günther, M., Höver, M., Kravtsov, S., Kuo, Y.-N., Lien, J., Madakumbura, G. D., Mankovich, N., Newman, M., Rader, J., Shi, J.-R., Shin, S.-I., and Varando, G.: Forced Component Estimation Statistical Method Intercomparison Project (ForceSMIP), ESS Open Archive, https://doi.org/10.22541/essoar.175003371.14843115/v1, 2025.

Wood, R. R., Lehner, F., Pendergrass, A. G., and Schlunegger, S.: Changes in precipitation variability across time scales in multiple global climate model large ensembles, Environmental Research Letters, 16, 084 022, https://doi.org/10.1088/1748-9326/ac10dd, 2021.

Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P., and Kruschke, T.: The SMHI Large Ensemble (SMHI-LENS) with EC-Earth3.3.1, Geoscientific Model Development, 14, 4781–4796, https://doi.org/10.5194/gmd-14-4781-2021, 2021.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, WIREs Climate Change, 2, 851–870, https://doi.org/https://doi.org/10.1002/wcc.147, 2011.

Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., Wang, Y., Dobrohotoff, P., Srbinovsky, J., Stevens, L., Vohralik, P., Mackallah, C., Sullivan, A., O'Farrell, S., and Druken, K.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.2288, 2019.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky, J.: The Australian Earth System Model: ACCESS-ESM1.5, JSHESS, 70, 193–214, https://doi.org/10.1071/ES19035, 2020.

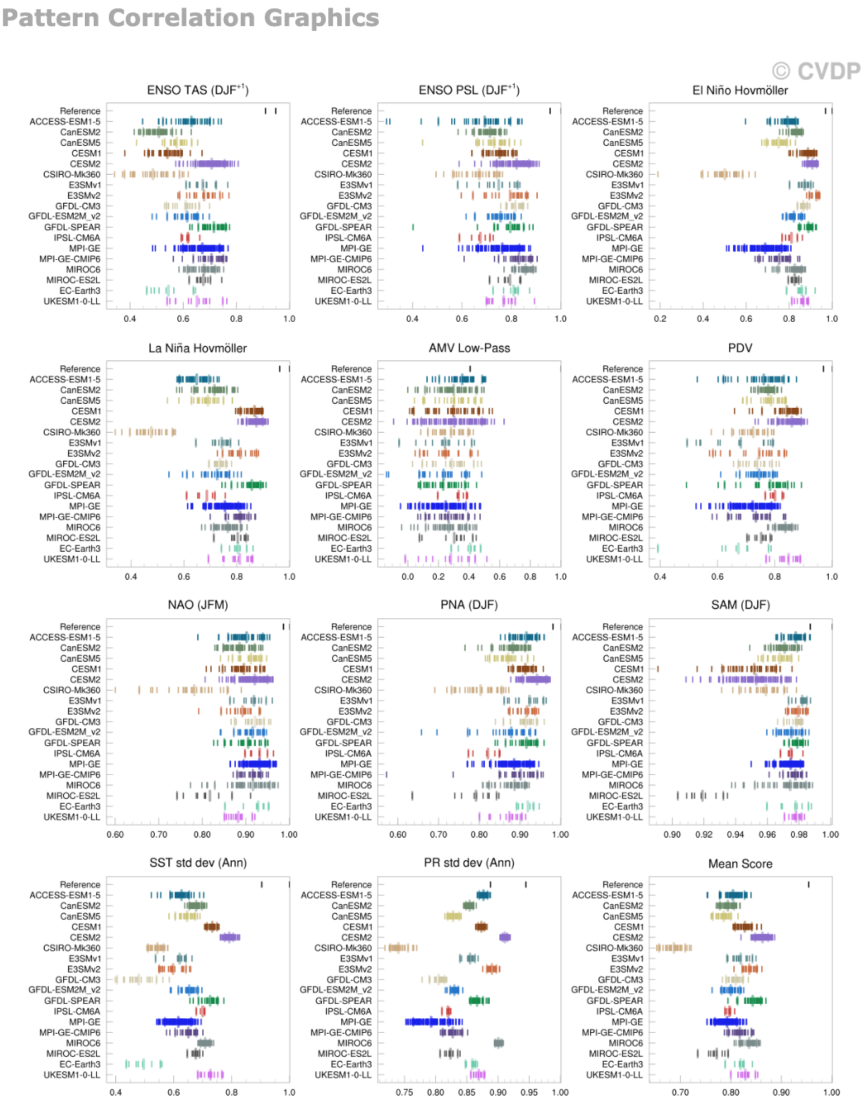
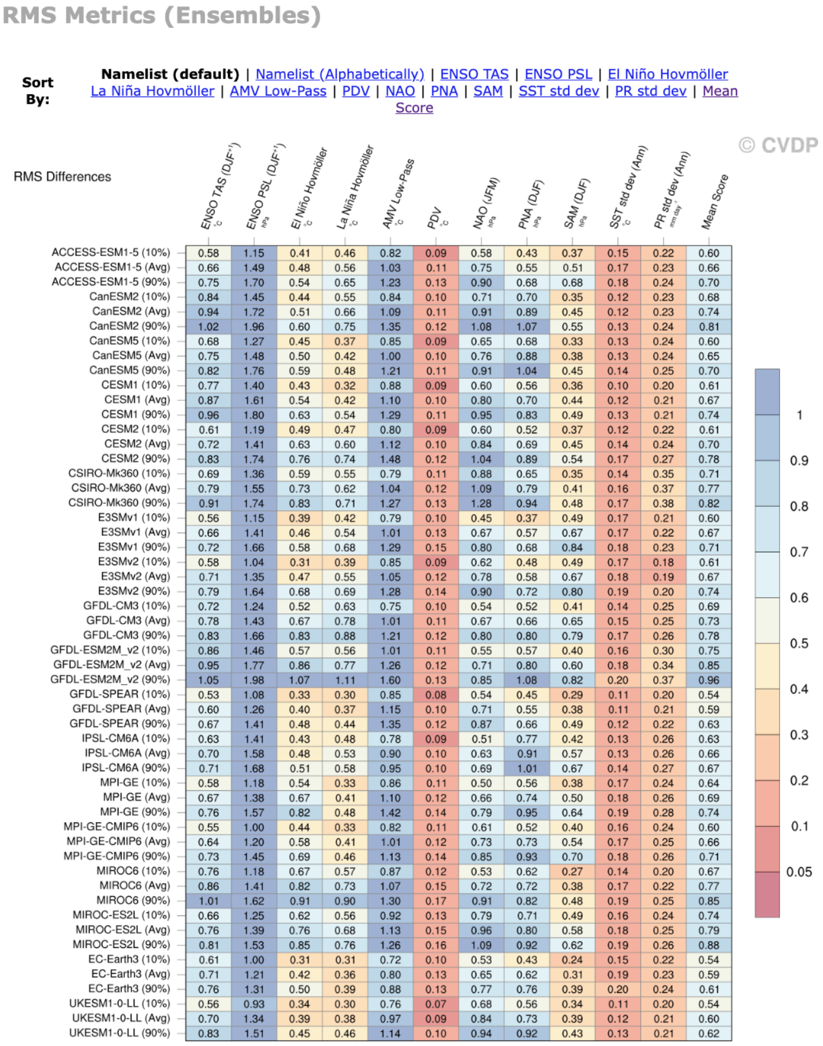


Figure 1. Pattern correlation of each ensemble member of all MMLEAv2 models with the first specified observational dataset (for each variable for each metric) is compared to every other specified observational dataset and every model simulation in this figure directly from the CVDPv6. Shown for multiple modes of variability, the standard deviation of sst and pr, and a mean score across all metrics used in the Figure (after applying a Fischer z-transform). Computations are completed over the period 1950-2022. Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Note that all of the metrics shown in Figs. 1 and 2 are spatial patterns. For example, "ENSO TAS (DJF+1)" is the spatial map of TAS anomalies in DJF+1 based on ENSO (e.g., El Nino minus La Nina) composites, "PR std dev (Ann)" is the map of PR standard deviation based on annual means, and "El Niño Hovmöller" is the Hovmöller diagram of El Niño composites of Equatorial Pacific SST anomalies over the longitude domain 120E-80W and time domain Jan year 0 – May year+2 (see Methodology and Definitions link in the CVDPv6). The following domains are used to compute the pattern correlations: Global for standard deviations, ENSO and PDV; 63S-65N for AMV; 120E-80W and Jan year0 – May year+2 for El Niño and La Niña Hovmöllers; 20N-90N for NAM/NAO; and 20S-90S for SAM.



RMS Metrics (Ensembles)

Sort By: **Namelist (default)** | Namelist (Alphabetically) | ENSO TAS | ENSO PSL | El Niño Hovmöller | La Niña Hovmöller | AMV Low-Pass | PDV | NAO | PNA | SAM | SST std dev | PR std dev | Mean Score

© CVDP

RMS Differences

	ENSO TAS (DJF^{+1}) °C	ENSO PSL (DJF^{+1}) hPa	El Niño Hovmöller °C	La Niña Hovmöller °C	AMV Low-Pass °C	PDV °C	NAO (JFM) hPa	PNA (DJF) hPa	SAM (DJF) hPa	SST std dev (Ann) °C	PR std dev (Ann) mm day^{-1}	Mean Score
ACCESS-ESM1-5 (10%)	0.58	1.15	0.41	0.46	0.82	0.09	0.58	0.43	0.37	0.15	0.22	0.60
ACCESS-ESM1-5 (Avg)	0.66	1.49	0.48	0.56	1.03	0.11	0.75	0.55	0.51	0.17	0.23	0.66
ACCESS-ESM1-5 (90%)	0.75	1.70	0.54	0.65	1.23	0.13	0.90	0.68	0.68	0.18	0.24	0.70
CanESM2 (10%)	0.84	1.45	0.44	0.55	0.84	0.10	0.71	0.70	0.35	0.12	0.23	0.68
CanESM2 (Avg)	0.94	1.72	0.51	0.66	1.09	0.11	0.91	0.89	0.45	0.12	0.23	0.74
CanESM2 (90%)	1.02	1.96	0.60	0.75	1.35	0.12	1.08	1.07	0.55	0.13	0.24	0.81
CanESM5 (10%)	0.68	1.27	0.45	0.37	0.85	0.09	0.65	0.68	0.33	0.13	0.24	0.60
CanESM5 (Avg)	0.75	1.48	0.50	0.42	1.00	0.10	0.76	0.88	0.38	0.13	0.24	0.65
CanESM5 (90%)	0.82	1.76	0.59	0.48	1.21	0.11	0.91	1.04	0.45	0.14	0.25	0.70
CESM1 (10%)	0.77	1.40	0.43	0.32	0.88	0.09	0.60	0.56	0.36	0.10	0.20	0.61
CESM1 (Avg)	0.87	1.61	0.54	0.42	1.10	0.10	0.80	0.70	0.44	0.12	0.21	0.67
CESM1 (90%)	0.96	1.80	0.63	0.54	1.29	0.11	0.95	0.83	0.49	0.13	0.21	0.74
CESM2 (10%)	0.61	1.19	0.49	0.47	0.80	0.09	0.60	0.52	0.37	0.12	0.22	0.61
CESM2 (Avg)	0.72	1.41	0.63	0.60	1.12	0.10	0.84	0.69	0.45	0.14	0.24	0.70
CESM2 (90%)	0.83	1.74	0.76	0.74	1.48	0.12	1.04	0.89	0.54	0.17	0.27	0.78
CSIRO-Mk360 (10%)	0.69	1.36	0.59	0.55	0.79	0.11	0.88	0.65	0.35	0.14	0.35	0.71
CSIRO-Mk360 (Avg)	0.79	1.55	0.73	0.62	1.04	0.12	1.09	0.79	0.41	0.16	0.37	0.77
CSIRO-Mk360 (90%)	0.91	1.74	0.83	0.71	1.27	0.13	1.28	0.94	0.48	0.17	0.38	0.82
E3SMv1 (10%)	0.56	1.15	0.39	0.42	0.79	0.10	0.45	0.37	0.49	0.17	0.21	0.60
E3SMv1 (Avg)	0.66	1.41	0.46	0.54	1.01	0.13	0.67	0.57	0.67	0.17	0.22	0.67
E3SMv1 (90%)	0.72	1.66	0.58	0.68	1.29	0.15	0.80	0.68	0.84	0.18	0.23	0.71
E3SMv2 (10%)	0.58	1.04	0.31	0.39	0.85	0.09	0.62	0.48	0.49	0.17	0.18	0.61
E3SMv2 (Avg)	0.71	1.35	0.47	0.55	1.05	0.12	0.78	0.58	0.67	0.18	0.19	0.67
E3SMv2 (90%)	0.79	1.64	0.68	0.69	1.28	0.14	0.90	0.72	0.80	0.19	0.20	0.74
GFDL-CM3 (10%)	0.72	1.24	0.52	0.63	0.75	0.10	0.54	0.52	0.41	0.14	0.25	0.69
GFDL-CM3 (Avg)	0.78	1.43	0.67	0.78	1.01	0.11	0.67	0.66	0.65	0.15	0.25	0.73
GFDL-CM3 (90%)	0.83	1.66	0.83	0.88	1.21	0.12	0.80	0.80	0.79	0.17	0.26	0.78
GFDL-ESM2M_v2 (10%)	0.86	1.46	0.57	0.56	1.01	0.11	0.55	0.57	0.40	0.16	0.30	0.75
GFDL-ESM2M_v2 (Avg)	0.95	1.77	0.86	0.77	1.26	0.12	0.71	0.80	0.60	0.18	0.34	0.85
GFDL-ESM2M_v2 (90%)	1.05	1.98	1.07	1.11	1.60	0.13	0.85	1.08	0.82	0.20	0.37	0.96
GFDL-SPEAR (10%)	0.53	1.08	0.33	0.30	0.85	0.08	0.54	0.45	0.29	0.11	0.20	0.54
GFDL-SPEAR (Avg)	0.60	1.26	0.40	0.37	1.15	0.10	0.71	0.55	0.38	0.11	0.21	0.59
GFDL-SPEAR (90%)	0.67	1.41	0.48	0.44	1.35	0.12	0.87	0.66	0.49	0.12	0.22	0.63
IPSL-CM6A (10%)	0.63	1.41	0.43	0.48	0.78	0.09	0.51	0.77	0.42	0.13	0.26	0.63
IPSL-CM6A (Avg)	0.70	1.58	0.48	0.53	0.90	0.10	0.63	0.91	0.57	0.13	0.26	0.66
IPSL-CM6A (90%)	0.71	1.68	0.51	0.58	0.95	0.10	0.69	1.01	0.67	0.14	0.27	0.67
MPI-GE (10%)	0.58	1.18	0.54	0.33	0.86	0.11	0.50	0.56	0.38	0.17	0.24	0.64
MPI-GE (Avg)	0.67	1.38	0.67	0.41	1.10	0.12	0.66	0.74	0.50	0.18	0.26	0.69
MPI-GE (90%)	0.76	1.57	0.82	0.48	1.42	0.14	0.79	0.95	0.64	0.19	0.28	0.74
MPI-GE-CMIP6 (10%)	0.55	1.00	0.44	0.33	0.82	0.11	0.61	0.52	0.40	0.16	0.24	0.60
MPI-GE-CMIP6 (Avg)	0.64	1.20	0.58	0.41	1.01	0.12	0.73	0.73	0.54	0.17	0.25	0.66
MPI-GE-CMIP6 (90%)	0.73	1.45	0.69	0.46	1.13	0.14	0.85	0.93	0.70	0.18	0.26	0.71
MIROC6 (10%)	0.76	1.18	0.67	0.57	0.87	0.12	0.53	0.62	0.27	0.14	0.20	0.67
MIROC6 (Avg)	0.86	1.41	0.82	0.73	1.07	0.15	0.72	0.72	0.38	0.17	0.22	0.77
MIROC6 (90%)	1.01	1.62	0.91	0.90	1.30	0.17	0.91	0.82	0.48	0.19	0.25	0.85
MIROC-ES2L (10%)	0.66	1.25	0.62	0.56	0.92	0.13	0.79	0.71	0.49	0.16	0.24	0.74
MIROC-ES2L (Avg)	0.76	1.39	0.76	0.68	1.13	0.15	0.96	0.80	0.58	0.18	0.25	0.79
MIROC-ES2L (90%)	0.81	1.53	0.85	0.76	1.26	0.16	1.09	0.92	0.62	0.19	0.26	0.88
EC-Earth3 (10%)	0.61	1.00	0.31	0.31	0.72	0.10	0.53	0.43	0.24	0.15	0.22	0.54
EC-Earth3 (Avg)	0.71	1.21	0.42	0.36	0.80	0.13	0.65	0.62	0.31	0.19	0.23	0.59
EC-Earth3 (90%)	0.76	1.31	0.50	0.39	0.88	0.13	0.77	0.76	0.39	0.20	0.24	0.61
UKESM1-0-LL (10%)	0.56	0.93	0.34	0.30	0.76	0.07	0.68	0.56	0.34	0.11	0.20	0.54
UKESM1-0-LL (Avg)	0.70	1.34	0.39	0.38	0.97	0.09	0.84	0.73	0.39	0.12	0.21	0.60
UKESM1-0-LL (90%)	0.83	1.51	0.45	0.46	1.14	0.10	0.94	0.92	0.43	0.13	0.21	0.62

Figure 2. RMS difference between each MMLEAv2 model ensemble average and observations as well as the 10th and 90th percentile of the RMS difference across the ensemble (shown in both colours and numbers in the table). Shown for multiple modes of variability, the standard deviation of sst and pr, and a mean score across all metrics used in the Figure (after normalizing each by the spatial RMS of the observed pattern to account for the different units of each variable). Computations are completed over the period 1950-2022. Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Area-weighted pattern correlations and RMS differences are calculated between observations and each model simulation (regridded to match the observational grid) for 11 climate metrics. The Total Score column shows the average of the 11 pattern correlations (Z-transformed) and RMS differences. The following domains are used to compute the RMS differences: Global for standard deviations, ENSO and PDV; 63S-65N for AMV; 120E-80W and Jan year0 – May year+2 for El Niño and La Niña Hovmöllers; 20N-90N for NAM/NAO; and 20S-90S for SAM..

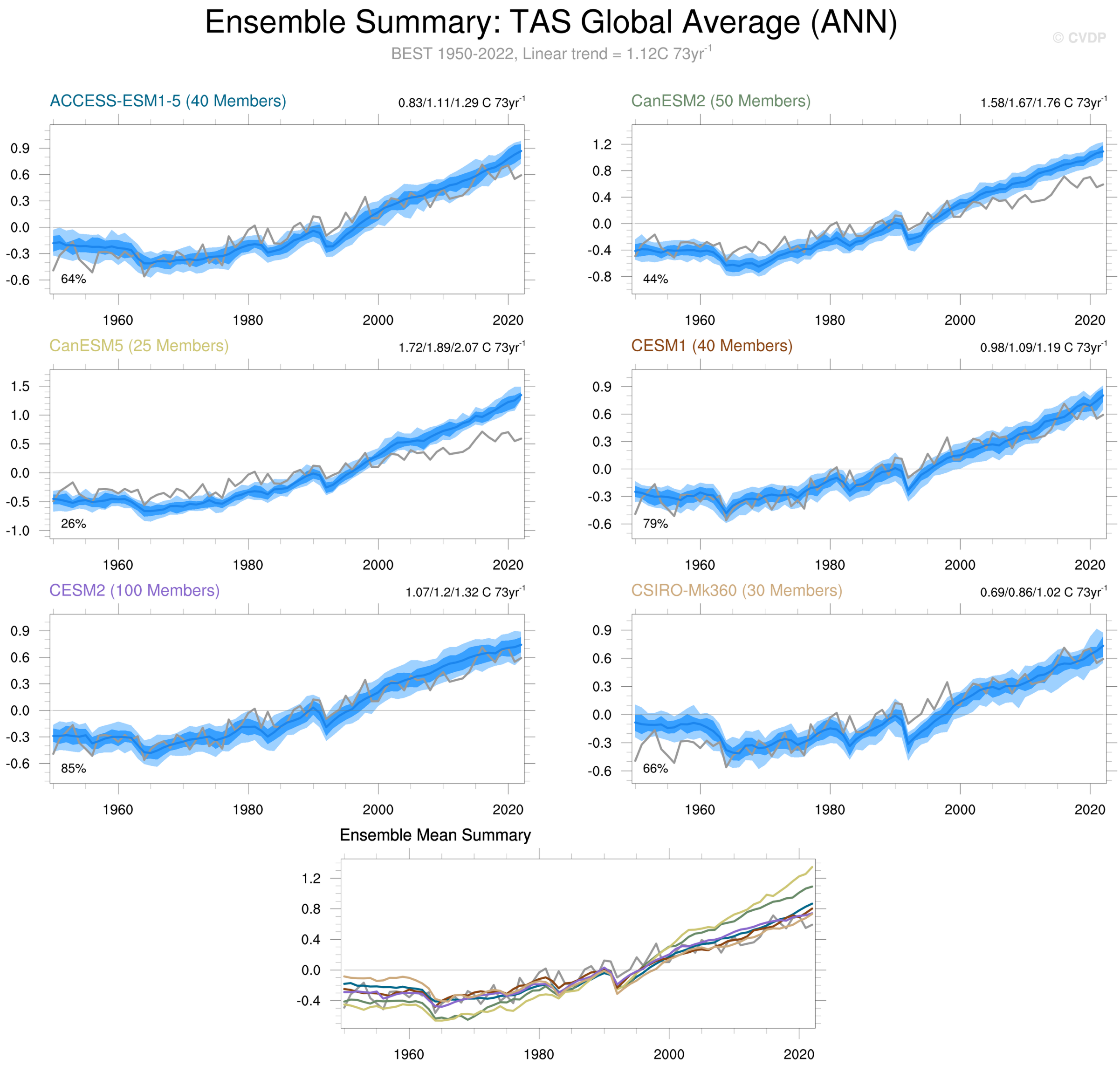


Figure 3. Global average surface air temperature anomalies for Set 1 computed annually over the period 1950-2022. The dark blue curve shows the model's ensemble mean timeseries, and the dark (light) blue shading around this curves depicts the 25th-75th (10th-90th) percentile spread across the ensemble members. Observations are shown in the thick gray curve with the dataset name and trend of the period written in gray under the Figure title. The last panel shows the ensemble mean of each LE (each model's line colour corresponds to the colour of each model's title in the previous panels of this figure) as well as observations. The 10th, 50th and 90th percentile values of the linear trends across the model are shown in the top right of each panel and the percentage value on the bottom left of each panel is the percentage of time that the observed value lies within the 10-90th percentile of the LE values. Here, each vertical bar denotes a different ensemble member, and the 10th, 50th and 90th percentile values are identified with taller bar. Set 2 & 3 can be found here: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set2_nonenone_1950-2022/tas_global_avg_ann.summary.png & https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set3_nonenone_1950-2022/tas_global_avg_ann.summary.png

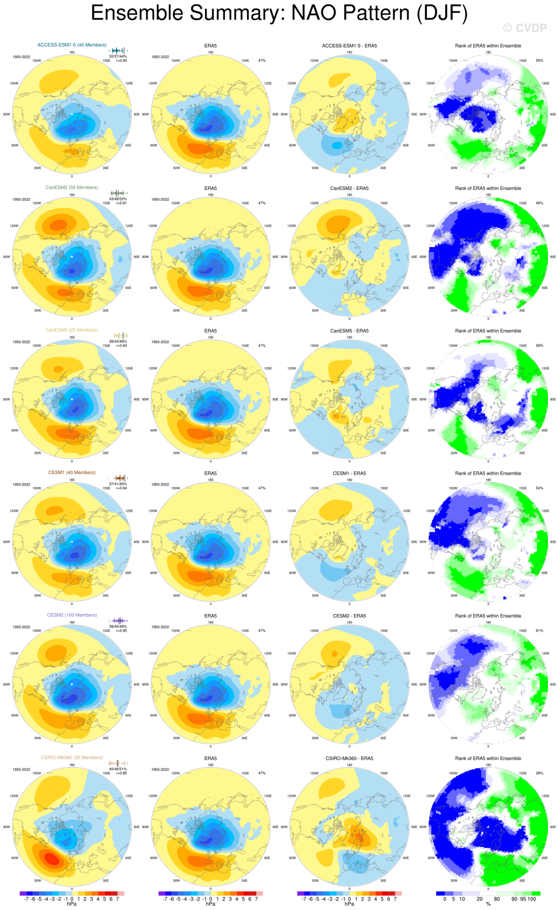


Figure 4. The North Atlantic Oscillation (NAO) Pattern in DJF for: left) ensemble mean of Set 1 models, middle left) observations, middle right) difference between the ensemble mean and observations and right) rank of observations within the ensemble spread. The percent variance explained (PVE) by the NAO over its native (EOF) domain is given in the subtitle at the top right of the regression map: the first, second and third values indicate the 10th, 50th and 90th percentile values across the ensemble, respectively. A graphical summary of the distribution of PVE values across the entire ensemble appears above the numerical values of PVE: each vertical bar denotes a different ensemble member, and the 10th, 50th and 90th percentile values are identified with taller bars. The observed PVE value is marked by a gray vertical bar. To quantify the degree of resemblance between the simulated and observed NAO regression maps, a pattern correlation (r) is computed between the observed NAO regression map and the ensemble average of the simulated NAO regression maps (e.g., maps in columns 1 and 2) over the domain shown. This r value is displayed at the upper right of each model panel, just below the range of PVE values White areas on the observed percentile rank maps indicate regions where the observed value lies within 10-90% of the model LE values, indicating the model is likely to be realistic. The value to the right of each rank map denotes the areal percentage of observed values lying within the 10th-90th percentile of the LE values. Computations are completed over the period 1950-2022. The NAO is the leading EOF in the region [20-80N, 90W-40E] following Hurrell and Deser (2009). Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Set 2 & 3 can be found here: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set2_quadrmEM_1950-2022/npo_pattern_djf.summary.png & https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set3_quadrmEM_1950-2022/nao_pattern_djf.summary.png

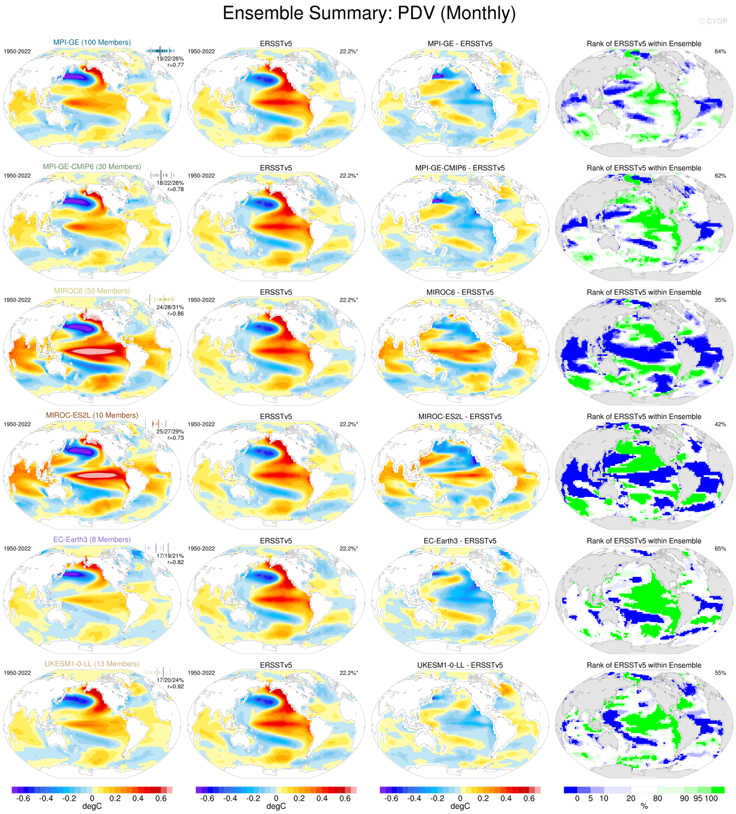


Figure 5. The Pacific Decadal Variability (PDV) Pattern in DJF for: left) ensemble mean of Set 3 models, middle left) observations, middle right) difference between the ensemble mean and observations and right) rank of observations within the ensemble spread. The percent variance explained (PVE) by the PDV over its native (EOF) domain is given in the subtitle at the top right of the regression map: the first, second and third values indicate the 10th, 50th and 90th percentile values across the ensemble, respectively. A graphical summary of the distribution of PVE values across the entire ensemble appears above the numerical values of PVE: each vertical bar denotes a different ensemble member, and the 10th, 50th and 90th percentile values are identified with taller bars. The observed PVE value is marked by a gray vertical bar. To quantify the degree of resemblance between the simulated and observed PDV regression maps, a pattern correlation (r) is computed between the observed PDV regression map and the ensemble average of the simulated PDV regression maps (e.g. maps in columns 1 and 2) over the domain shown. This r value is displayed at the upper right of each model panel, just below the range of PVE values White areas on the observed percentile rank maps indicate regions where the observed value lies within 20-80% of the model LE values, indicating the model is likely to be realistic. The value to the right of each rank map denotes the areal percentage of observed values lying within the 10th-90th percentile of the LE values. Computations are completed over the period 1950-2022. The PDV Index is defined as the standardized principal component (PC) timeseries associated with the leading Empirical Orthogonal Function (EOF) of area-weighted monthly SST anomalies over the North Pacific region [20-70N, 110E-100W] minus the global mean [60N-60S] following Mantua et al. (1997). Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Set 1 & 2 can be found here: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set1_quadrmEM_1950-2022/pdv.summary.png & https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set3_quadrmEM_1950-2022/pdv.summary.png

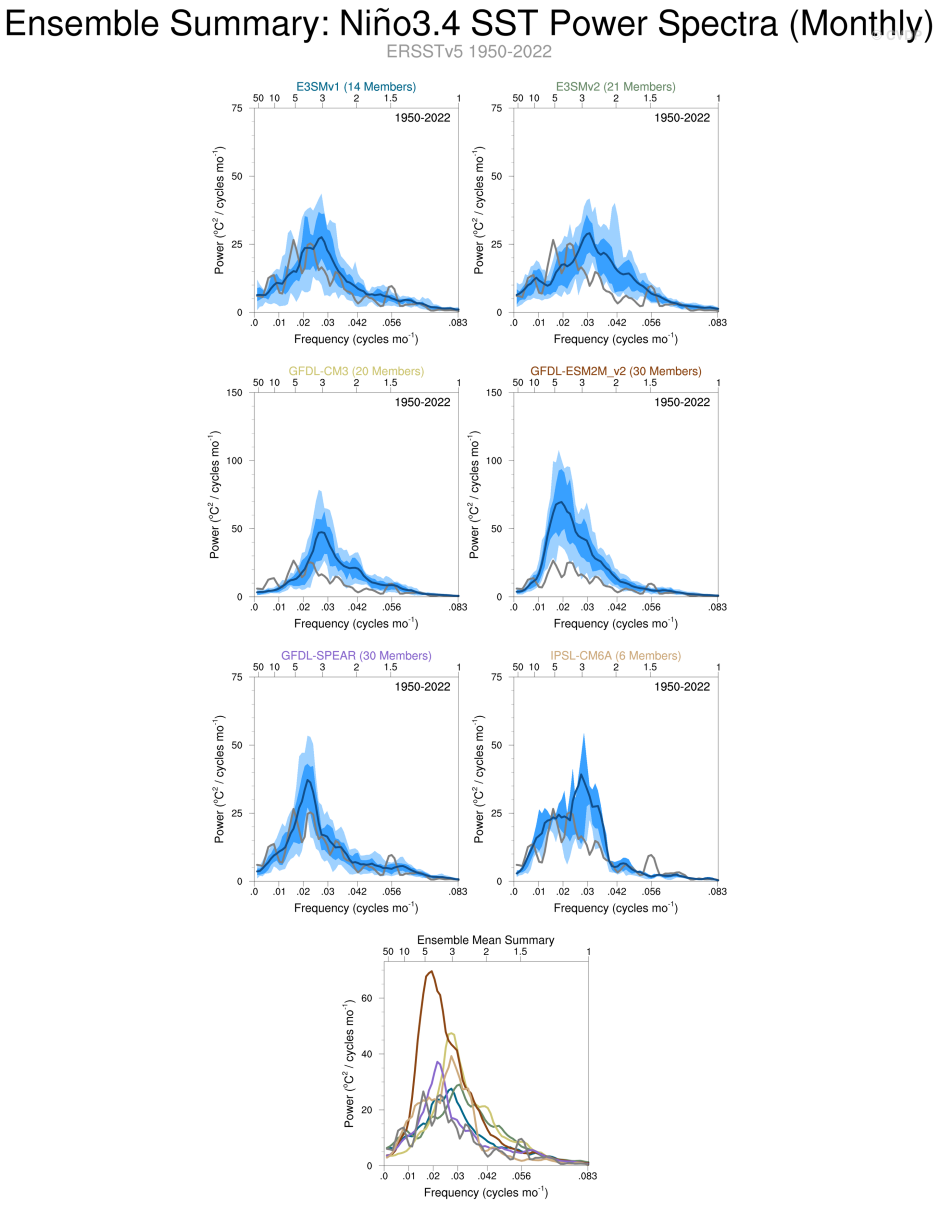


Figure 6. ENSO (nino3.4) monthly power spectrum for Set 2 models. The dark blue curve shows the model's ensemble mean power spectrum, and the dark (light) blue shading around this curves depicts the 25th-75th (10th-90th) percentile spread across the ensemble members and the grey/brown line is the observations. The bottom right shows the ensemble mean from each model and the observations on a single panel. Computations are completed over the period 1950-2022. Power spectra are computed in variance preserving format from the linearly detrended December Nino3.4 SST Index (SST anomalies averaged over the region 5N-5S, 170W-120W). Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Set 1 & 3 can be found here: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set1_quadrmEM_1950-2022/nino34.powspec.summary.png & https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set3_quadrmEM_1950-2022/nino34.powspec.summary.png

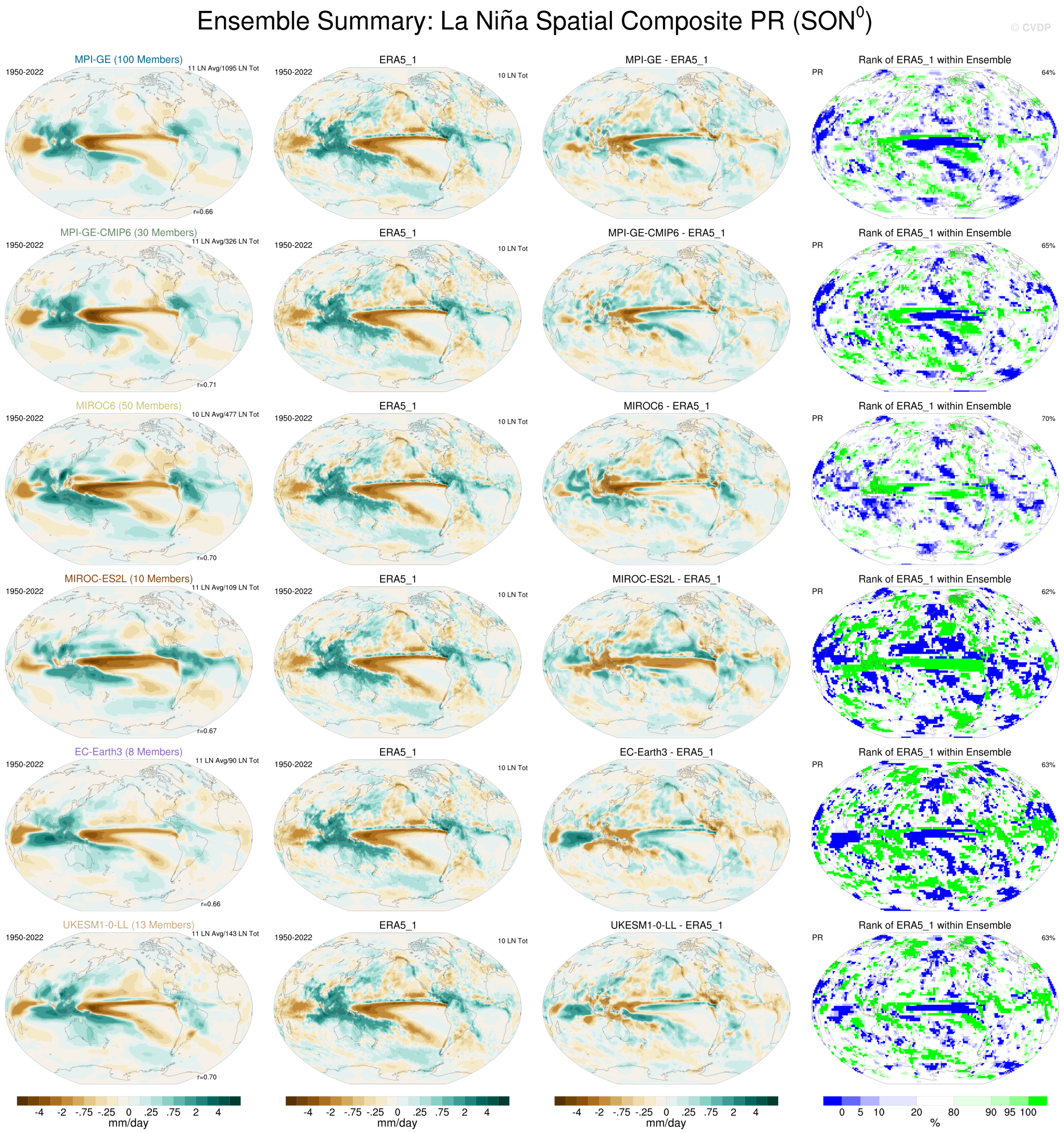


Figure 7. Composite of precipitation in SON for La Niña events for: left) ensemble mean of Set 3 models, middle left) observations, middle right) difference between the ensemble mean and observations and right) rank of observations within the ensemble spread. The top right of the rightmost panels shown the percentage of the map that is white, where observations are considered to be within the ensemble spread. Computations are completed over the period 1950-2022. The pattern correlations are displayed at the lower right of each model panel. The number of events that go into each spatial composite are displayed at the upper right of each panel (given as an average per ensemble member and as a total over all ensemble members). The value to the right of each rank map denotes the areal percentage of observed values lying within the 10th-90th percentile of the LE values. Observations are detrended using a quadratic fit, while LEs are detrended by removing the ensemble mean. Set 1 & 2 can be found here: https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set1_quadrmEM_1950-2022/nino34.spatialcomp.lanina.pr.summary.son0.png & https://webext.cgd.ucar.edu/Multi-Case/MMLEA_v2/MMLEA_Set2_quadrmEM_1950-2022/nino34.spatialcomp.lanina.pr.summary.son0.png

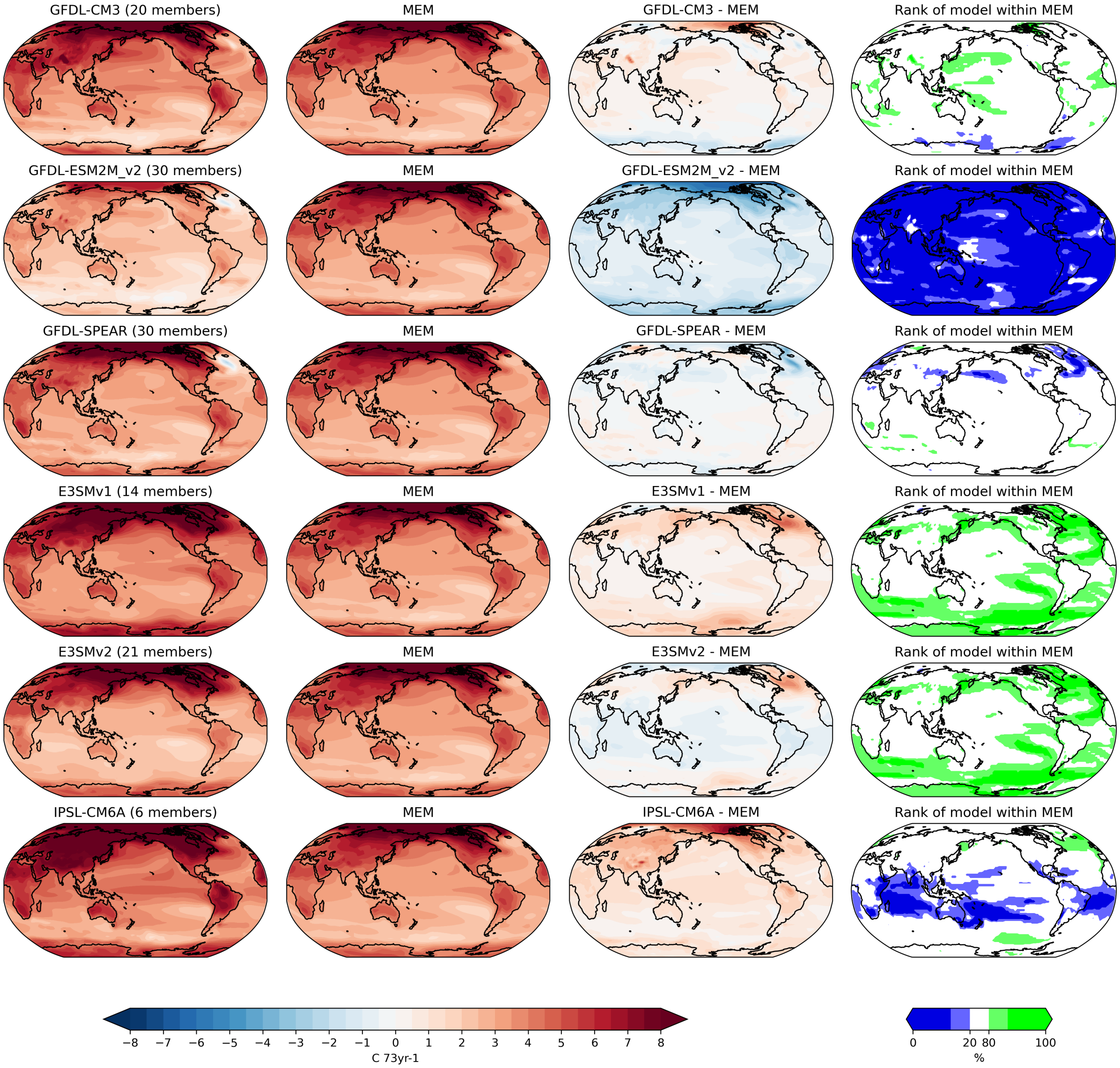


Figure 8. Annual surface air temperature trend from 2027-2099 for: left) ensemble mean of Set 2 models, middle left) multi-ensemble mean (MEM), middle right) difference between the ensemble mean and the MEM and right) rank of the model ensemble mean within the MEM. All analysis is completed on RCP8.5 and SSP5-8.5 except for GISS-E2-G,CESM2 and UKESM1-0-LL where SSP3-7.0 is used.

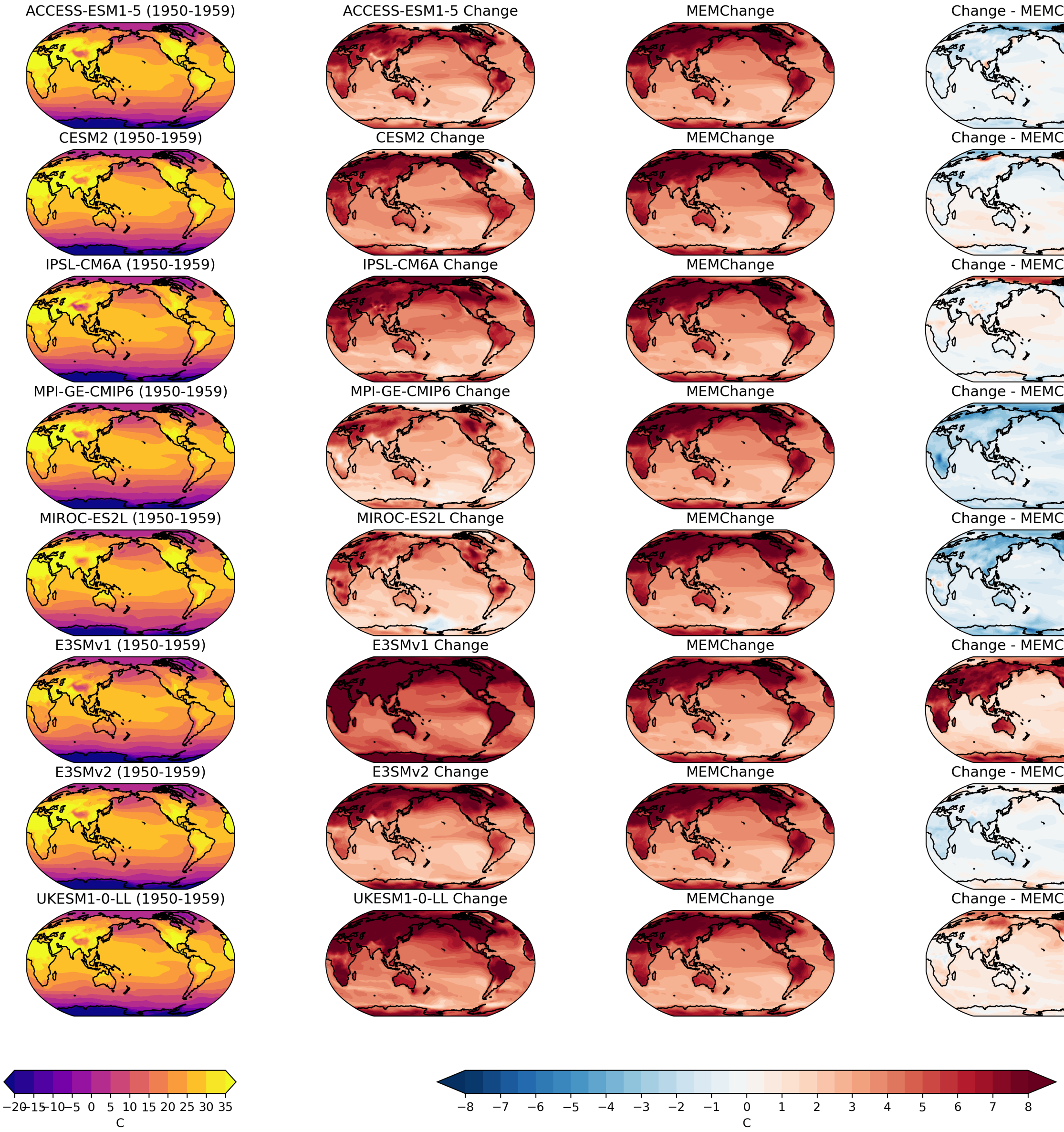


Figure 9. Monthly maximum of the daily maximum temperature (TXx) averaged over June-July-August in the 8 models that have data for the SSP370 future scenario. left) Individual model ensemble mean for the period 1950-1959, middle left) the change in 2090-2099 as compared to 1950-1958, middle right) the change in 2090-2099 as compared to 1950-1959 in the multi-ensemble mean (MEM) of the 8 models shown on the plot, right) change in the single model ensemble mean minus the MEM change.