Assessing Climate Variability and Change in Model Large Ensembles:

A User's Guide to the "Climate Variability Diagnostics Package for Large Ensembles"

Adam S. Phillips, Clara Deser, John Fasullo, David P. Schneider and Isla R. Simpson

Climate Analysis Section, NCAR

16 October 2020

Author contact information:

Adam Phillips asphilli@ucar.edu

Clara Deser cdeser@ucar.edu

Table of Contents

1. Preface

Welcome to the world of Large Ensembles! This document serves two purposes. One is to introduce the reader to the concept of "Large Ensembles" and why they are crucial for understanding the observed climate record, evaluating climate models, and providing society with information on the risks of future climate change due to human activities. The second is as a user's manual for the "Climate Variability Diagnostics Package for Large Ensembles", an automated analysis tool and data repository that facilitates exploration of modes of climate variability and change in model large ensembles and observations. We hope that these resources promote new discoveries by students, educators and researchers alike. We welcome your feedback on this document and the analysis package itself, including suggestions for additional variables, metrics and graphical displays.

2. Introduction

The climate system is highly variable on all space and time scales. Apart from external influences associated with natural (e.g., volcanic eruptions and orbital cycles that impact insolation) and anthropogenic (e.g., greenhouse gases and sulfate/biomass burning aerosols) factors, this variability stems from processes intrinsic to the climate system. For example, non-linear dynamics of the atmosphere spontaneously generates day-to-day weather events such as storms and atmospheric rivers, as well as longer-lived (week-to-week and month-to-month) changes in jet-stream level winds and associated storm activity around the world. Non-linear processes also operate in the ocean and give rise to slower (year-to-year and decade-to-decade) variations that influence the atmosphere and climate. Finally, non-linear interactions between the atmosphere and ocean produce climate fluctuations on seasonal and interannual time scales,

including the well-known El Niño – Southern Oscillation (ENSO) phenomenon. The chaotic non-linear nature of the climate system means that predictability of such intrinsic fluctuations is limited.

Internally-generated weather and climate variations co-exist with externally-driven influences on the climate system, whether natural or human-caused. Indeed, their prevalence obscures identification of anthropogenic contributions to observed climate anomalies, and necessitates probabilistic projections of future climate change. Intrinsic weather and climate fluctuations may, in turn, be influenced by external forcing. For example, there is evidence that volcanic eruptions may trigger weak ENSO events and that rising greenhouse gases may increase the volatility of precipitation, exacerbating drought and flood extremes.

A major challenge for climate researchers is to construct numerical models of the Earth's climate system that include realistic representations of both internally-generated and externally-driven sources of climate variability and change. Evaluation of such models is of key importance and presents its own set of challenges, including those related to computational and data storage requirements as well as limitations imposed by the instrumental record. Automated tools have been developed to facilitate climate model assessment such as "ESMValTool" (Eyring et al., 2016), which is comprised of a collection of analysis packages developed at various climate modeling centers around the world. One of the component packages of ESMValTool is the National Center for Atmospheric Research (NCAR) "Climate Variability Diagnostics Package" (Phillips et al., 2014) upon which the "Climate Variability Diagnostics Package for Large Ensembles" is based.

3. The Importance of Large Ensembles

A new thrust in climate modeling is to run a large number of simulations with the same model and the same radiative forcing protocol but varying the initial conditions, sometimes by

extremely small amounts on the order of the model's numerical round-off error (e.g., 10^{-14} K for atmospheric temperature). The differences in initial conditions serve to create ensemble spread once the memory of the initial state is lost, typically within a month for the atmosphere and a few years to a decade for the ocean. The resulting ensemble spread is thus solely due to unpredictable internally-generated climate variability (aka, the "butterfly effect" in chaos theory). Because the temporal sequences of internal variability differ across the simulations once the memory of the initial conditions is lost, one can average the members ("members" will be used synonymously with "simulations") together to estimate the forced response at each time step, although a sufficient ensemble size is needed to do so accurately (Milinski et al., 2020). It is important to note that each simulation in the ensemble contains a common forced response superimposed upon a different sequence of internal variability.

Such "initial-condition Large Ensembles" (LEs) have proven to be enormously useful for separating internal variability and forced climate change on regional and decadal scales (e.g., Deser et al., 2012; Kay et al., 2015; Maher et al., 2019; Deser et al. 2020; Lehner et al., 2020). They have also been used to assess externally-forced changes in the characteristics of internal variability, including extreme events for which large sample sizes are crucial. Additionally, they have served as methodological testbeds for evaluating approaches to detection and attribution of anthropogenic climate change in the (single) observational record (e.g., Deser et al., 2016; Barnes et al., 2019; Sippel et al., 2019; Santer et al. 2019; Bonfils et al., 2019; Wills et al., 2020). Until the advent of LEs, it was problematic to identify the sources of model differences in the Coupled Model Intercomparison Project (CMIP) archives due to the limited number of simulations (generally 1-3) for each model (i.e., structural uncertainty was confounded with uncertainty due to internal variability) (see for example Deser et al., 2012 and Lehner et al., 2020).

Just as in a model LE, the sequence of internal variability in the real world could have unfolded differently. That is, the observational record traces only one of many possible climate histories that could have happened under the same external forcing (i.e., volcanoes, solar and orbital cycles, and anthropogenic emissions). For example, El Niño and La Niña events could have occurred in a different set of years, or interannual fluctuations of the "North Atlantic Oscillation" could have occurred with an alternate chronology. This concept is sometimes referred to as the "Theory of Parallel Climate Realizations" (Tel et al., 2019) or the notion of "Contingency" (Gould, 1989). The implications of this concept are enormous. For one, it means that a single model simulation of the historical period need not match the observed record, even if the model is "perfect" in its representation of statistical characteristics of the real world's climate. However, the statistical characteristics of the model's internal variability must match those of the real world, taking into account the limited sampling in the observational record due to the limited ensemble size (one) and limited duration and spatial coverage of the measurements. Observational uncertainty due to instrument error, data inhomogeneities, and application of optimal interpolation and reanalysis methods must also be taken into account (see the Climate Data Guide https://climatedataguide.ucar.edu/ for expert guidance on observational uncertainty). Further, the spread across a model LE must encompass the observational record (subject to the above observational uncertainties) for the model to be credible, provided there are enough members to adequately span the range of possible sequences of internal variability. However, this is not a sufficient criterion, since a model with unrealistically large internal variability may encompass the observational record for the wrong reason [see Deser et al. (2017b) for an expanded discussion of this point with application to ENSO teleconnections]. These considerations emphasize the importance of evaluating both the amplitudes and patterns of internal variability in models.

4. The CVDP

The NCAR "Climate Variability Diagnostics Package" (CVDP) is an analysis tool that documents the major modes of climate variability as well as long-term trends and other metrics in models and observations based on a variety of fields including sea surface temperature, surface air temperature, sea level pressure, precipitation and sea ice concentration (http://www.cesm.ucar.edu/working_groups/CVC/cvdp/). 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. A recent application of the CVDP to the CMIP 3, 5 and 6 archives highlighted general improvements in the simulation of the major modes of climate variability across the three generations of CMIP models (Fasullo et al., 2020). The CVDP has served the community as both a research and teaching resource.

One way that the CVDP facilitates model inter-comparison and observational uncertainty is by displaying graphical information from all models and observational data sets on a single page. In addition to graphical output (in png format), the CVDP produces netcdf files for each diagnostic and saves both the png and netcdf files to a web-based data repository for later access and subsequent analysis. The CVDP is coded with the open source "NCAR Command Language (NCL)", but the user does not need to know NCL to apply it or to make use of its output. All calculations performed in the CVDP are fully documented with references to the literature in the "Methodology" link. To apply the CVDP, the user specifies a suite of observational data sets and model simulations in the "Input Namelist". Multiple observational data sets can be specified for the same quantity (for example, sea surface temperature), and multiple time periods can also be

specified if desired (even for the same data set). In this way, the sensitivity of the results to the choice of observational data set and time period can be investigated. The user can also choose to pre-process the model and observational data before applying the CVDP. For example, the user may wish to detrend or low-pass filter the data before supplying them to the CVDP.

5. The CVDP for Large Ensembles

To take advantage of the unique information provided by LEs, we have developed a new version of the CVDP called the "CVDP-LE". This resource includes the same diagnostics as the original CVDP, but computes and displays the information in ways that are specifically tailored to LEs. For example, the ensemble-mean (an estimate of the forced response) and the ensemble spread due to internal variability are computed and displayed separately. Like its predecessor, the CVDP-LE operates on a user-specified set of model simulations and observational datasets, and saves the output (.png graphical displays and netcdf data files) to a data repository for later access and further analysis. Unique to the CVDP-LE is the ability to view the output from two perspectives: "Individual Member" and "Ensemble Summary". The former is analogous to the format used in the CVDP; the latter is specific to the CVDP-LE.

The rest of this document is organized as follows. Section 5a provides an overview of how the diagnostics in the CVDP-LE are arranged. Section 5b presents a short discussion of how to deal with observational uncertainty. Section 5c provides guidance on how to interpret the information shown in the "Ensemble Summary" view, including spatial patterns, timeseries and derived quantities, and summary metrics tables and graphics. Section 5d gives a brief description of the "Individual Member" view. Section 6 "Tips for applying the CVDP-LE" includes some principles for "best practices" in LE analysis including caveats and potential pitfalls, along with

ideas for novel ways to pre-process the input data for added insights. Instructions on how to download the CVDP-LE code and run the package may be found at http://www.cesm.ucar.edu/working_groups/CVC/cvdp-le/.

5a. Organization of diagnostics

The diagnostics computed in the CVDP-LE are organized in a similar way to those in the CVDP. They begin with general quantities (climatological averages, interannual standard deviations, and linear trends; defined below) displayed in the form of global maps for each variable and season. The variable names are abbreviated following CMIP convention as follows: SST for sea surface temperature, TAS for 2m Air Temperature, PSL for barometric sea level pressure, and SIC for sea ice concentration. The seasons are 3-month averages for: December-February (DJF), January-March (JFM), March-May (MAM), June-August (JJA), July-September (JAS). Annual means (January-December averages) are also shown. Next are the major modes of coupled (ocean-atmosphere) climate variability: El Nino – Southern Oscillation (ENSO), Atlantic Multidecadal Variability (AMV), Pacific Decadal Variability (PDV), and the Atlantic Multidecadal Overturning Circulation (AMOC). These are followed by the dominant modes (8 in total) of large-scale atmospheric circulation variability. The final set of diagnostics is devoted to time series, including globally-averaged quantities, regional indices and a variety of sea ice extent measures. All of the diagnostics are defined in detail below, along with a description of the various metrics used to characterize the modes of variability.

General Quantities

a) Climatological averages: averages over the time period specified by the user.

b) Interannual standard deviations: temporal standard deviations based on linearly detrended data during the time period specified by the user.

c) Linear trends: linear least-squares best-fit trends over the time period specific by the user.

Coupled Modes of Variability

a) ENSO

The CVDP-LE provides an extensive set of diagnostics on ENSO given its central importance to the global climate system. These diagnostics are grouped into two categories: spatial patterns and temporal characteristics. The spatial patterns are based on compositing all El Niño events and all La Niña events using a +1 and -1 standard deviation threshold of the linearly detrended December Nino3.4 SST Index (SST anomalies averaged over the region 5N-5S, 170W-120W) to define events, respectively, following previous studies (e.g., Deser et al., 2010). Global maps of linearly detrended SST, TAS, PSL and PR anomalies based on composite El Niño events, composite La Niña events, and their difference are shown for each season, along with Hovmöller diagrams of the space-time composite evolution of equatorial Pacific SST anomalies (averaged between 3N-3S) for the El Niño and La Niña composites. The seasons are labeled relative to December of the year used to define events (denoted year 0). Thus, JJA^0 and SON^0 correspond to year 0, and DJF^1 and MAM^1 correspond to year +1. Metrics of the temporal characteristics, based on the detrended monthly Nino3.4 SST Index, include standard deviations by month, power spectra (in variance preserving format), Morlet wavelets (Torrence and Compo, 1998), lag-autocorrelation curves, and time-evolving standard deviations based on running 30-year windows. The raw (i.e., un-detrended) Nino3.4 SST Index timeseries are also provided for context (note that all of the other ENSO metrics are based on linearly detrended data).

b) AMV

The AMV Index is defined as monthly SST anomalies averaged over the North Atlantic region [0-60N, 80W-0W] minus the global mean [60N-60S] following Trenberth and Shea (2006). Both unfiltered and low-pass filtered (based on 10-year running means) versions of the AMV Index timeseries are provided, along with the power spectrum (in variance preserving format) of the unfiltered timeseries. Global regression maps of monthly SST, TAS and PR anomalies onto the AMV Index timeseries are also provided for the unfiltered ("Regr") and low-pass filtered ("Regr LP") versions of the data.

c) PDV

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). The diagnostics provided for PDV are the same as those for AMV above.

d) Alternate definitions of AMV and PDV

The rationale for subtracting the global mean SST in the definitions of the AMV and PDV indices is to reduce any potential influence of external radiative forcing such as GHG. However, if the pattern associated with global warming projects strongly onto the patterns associated with AMV and PDV, then this subtraction will be inadequate for its intended purpose. For this reason, we have included alternative indices of AMV and PDV (termed AMV' and PDV') which isolate

the internal component of these modes by subtracting the ensemble-mean value from each ensemble member at each time step for a given model LE (Deser et al., in preparation). This subtraction is performed for the Index timeseries and all of the fields used in the regression analyses.

e) AMOC

The Atlantic Meridional Overturning Circulation (AMOC) Index is defined as the standardized PC timeseries associated with the leading EOF of area-weighted annual mean oceanic meridional mass transport (Sv) in the Atlantic sector from 33°S to 90°N over the depth range 0 – 6km following Danabasoglu et al. (2012). This Index is then low-pass filtered with a 15-point Lanczos filter. Both the timeseries and power spectrum (in variance preserving format) of the low-pass filtered AMOC Index are displayed. The spatial patterns of oceanic meridional mass transport (Sv) associated with AMOC are shown as a function of latitude and depth for the climatological mean, interannual standard deviation, and regression patterns onto the AMOC Index. In addition, global regression maps (at zero lag) of annual low-pass filtered SST and TAS (in degrees Celsius) anomalies associated with a one standard deviation departure of the AMOC Index are displayed. Finally, lead/lag correlation curves from -15 years to +15 years between the low-pass filtered AMOC and AMV timeseries are shown.

Atmospheric Modes of Variability

The following modes of large-scale atmospheric circulation variability are included based on their regional importance; all are defined using monthly area-weighted PSL anomalies (departures from the long term means for each month):

- Southern Oscillation (SO): Difference between the Indian Ocean/Western Pacific [70-170E] and the Central/Eastern Pacific [160W-60W] averaged over the latitude band 30S-0 (see Trenberth and Caron, 2000).
- Northern Annular Mode (NAM): leading EOF north of 20N (Thompson et al., 2000)
- North Atlantic Oscillation (NAO): leading EOF in the region [20-80N, 90W-40E] following Hurrell and Deser (2009)
- Southern Annular Mode (SAM): leading EOF south of 20S (Thompson et al., 2000)
- Pacific - North American Pattern (PNA): leading EOF in the region [20-85N, 120E-120W]
- North Pacific Oscillation (NPO): second EOF in the region [20-85N, 120E-120W]
- Pacific – South American Pattern Mode 1 (PSA1): second EOF south of 20S (Mo and Higgins, 1998)
- Pacific – South American Pattern Mode 2 (PSA2): third EOF south of 20S (Mo and Higgins, 1998)

For the EOF-based modes, the Index timeseries are the associated PCs. All Index timeseries are standardized (i.e., divided by their standard deviations). The spatial patterns of the modes are displayed as hemispheric (global in the case of the SO) PSL regression maps upon the standardized Index timeseries. The modes are computed separately for each season and the annual mean. Unlike SST, the externally-forced component of these PSL-based modes is generally small compared to the internally-generated component in models (e.g., Deser et al., 2012), and hence for simplicity we have chosen to not detrend the data before computing them. Note, however, that the user can

detrend the PSL data (or, alternatively, subtract the ensemble-mean PSL values from each ensemble member at each time step) before applying the CVDP if this is a concern.

Timeseries

a) Global: Area-weighted global averages of SST, TAS and PR anomalies for each season and the annual mean. Land-only PR is also included, as many observational data sets before the satellite-era are land-only.

b) Regional: Area-weighted monthly SST anomaly timeseries as follows (see the cited references for information on their climatic significance).

- Tropical North Atlantic (5N-23N, 15-60W, Enfield et al. 1999)
- Tropical South Atlantic (0-20S, 30W-10E, Enfield et al. 1999)
- Atlantic Meridional Mode (Doi et al., 2009)
- Atlantic Niño (3N-3S, 20W-0E; Zebiak, 1993)
- North Atlantic SST (0-60N, 80W-0W)
- Tropical Indian Ocean (15S-15N, 40-110E)
- Indian Ocean Dipole [10S-10N, 50-70E] minus [0-10S, 90-110E] (Saji et al., 1999)
- Niño1+2 (0-10S, 80-90W; Rasmusson and Carpenter, 1982)
- Niño3 (5N-5S, 120-170W; Rasmusson and Carpenter, 1982)
- Niño4 (5N-5S, 160E-150W; Rasmusson and Carpenter, 1982),
- North Pacific Meridional Mode (Chiang and Vimont, 2004; Amaya 2020),
- South Pacific Meridional Mode (Zhang et al., 2014; Amaya 2020),
- Southern Ocean (50-70S, 0-360E)

The North Pacific PSL Index (NPI; Trenberth and Hurrell, 1994) is also included. This Index is defined as the area-weighted PSL anomaly averaged over the region [30-65N, 160E-140W] and averaged over the months December-March.

c) Sea ice extent: Monthly, seasonal and annual averages of sea ice extent in the Northern and Southern Hemispheres (NH and SH, respectively). Sea ice extent is defined as the area of ocean with at least 15% sea ice. Monthly anomalies and the monthly climatology of sea ice extent are also included.

5b. Treatment of Observational Uncertainty

It is up to the user to decide which observational data sets to use for evaluating the model LEs. A list of commonly used observational products is provided in the CVDP (http://www.cesm.ucar.edu/working_groups/CVC/cvdp-le/observations.html), along with links to their sources and information on strengths, limitations and applicability from the Climate Data Guide. If multiple observational datasets are specified for a given variable (for example, SST) to assess observational uncertainty, the first one in the "Input Namelist" is the one used for model evaluation in the CVDP-LE. Thus, it is good practice to provide a dataset that you consider your "best guess" as the first one in the list. However, you can simply re-run the CVDP-LE with a different observational product listed first to assess sensitivities to observational uncertainty. Further, for each diagnostic, the "Individual Member" view contains results for all of the observational data sets specified by the user in a single graphic, providing a direct comparison and facilitating assessment of observational uncertainty. Finally, the "Metrics Graphics" and "Metrics Individual Tables" (described in Section 5c) includes comparisons of all the different observational

products for 11 key metrics of climate variability, providing context for model performance with respect to observational uncertainty.

5c. "Ensemble Summary" View

As mentioned above, the output from the CVDP-LE can be viewed from two perspectives: "Individual Member" and "Ensemble Summary". The former presents the diagnostics for each model simulation individually (analogous to the information provided by the original CVDP) and does not leverage the unique information inherent to LEs. In contrast, the latter provides information on the ensemble-mean and ensemble-spread of each model LE, in addition to a ranking of observations relative to the model's ensemble spread. These metrics are provided for each diagnostic (i.e., climatological averages, interannual standard deviations, linear trends, modes of variability, and timeseries). The metrics for a particular diagnostic quantity in all model LEs are displayed graphically on a single page, facilitating model inter-comparison and performance. Below, we provide some examples of diagnostics from the "Ensemble Summary" view and explain how to interpret the information provided. Our examples are drawn from an application of the CVDP-LE to the "Multi-Model Large Ensemble Archive" (MMLEA; Deser et al., 2020) over the time period 1950-2018; the complete set of CVDP-LE diagnostics from this comparison are available at http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/. The MMLEA is a collection of data from 7 CMIP5-class model LEs assembled by the US CLIVAR Working Group on Large Ensembles (https://usclivar.org/working-groups/large-ensemble-working-group). This archive and the potential insights that it affords are described in Deser et al. (2020).

Interpreting Spatial Patterns

Here, we discuss how to make use of the spatial information provided for each diagnostic in the "Ensemble Summary" view of the CVDP-LE. For illustration purposes, we use the example of the DJF NAO during 1950-2018, see: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nao_pattern_djf.summary.png. For convenience, this plot is reproduced in Fig. 1 at the end of this document. The left-hand column of maps shows the spatial pattern of the NAO for each of the 7 model LEs arranged in the order in which they were specified in the "Input Namelist" (alphabetical in this case). The title above each map identifies the model by name and the number of ensemble members used in the analysis; this information is color-coded by model for added visibility (the same color-coding is used in the "Individual Member" maps, where it is particularly useful for distinguishing the different models). The subtitle at the top left of each map identifies the period of record analyzed. For example, the map at the top left of Fig. 1 shows the NAO pattern simulated in the CCCma LE using 40 members during 1950-2018.

As discussed previously, the NAO is computed as the leading EOF of area-weighted DJF PSL anomalies over the North Atlantic/European domain (20 - 80°N, 90°W - 40°E) following Hurrell and Deser (2009). The map displays the linear regression of DJF PSL anomalies at each grid box north of 20°N onto the standardized principal component (PC) timeseries, providing a hemispheric view of the expression of the regional NAO. The "ensemble summary" view of the NAO pattern is obtained by performing a separate EOF (and regression) calculation for each ensemble member and then averaging the resulting regression maps over all available ensemble members. Note that the individual regression maps are viewable from the "Individual Member" link: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nao_pattern_djf.indmem.png.

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. For example, in CCCma these values are: 42%, 48% and 53%. [Throughout the CVDP-LE, we shall use the 10^{th} - 90^{th} percentile range as a measure of statistical likelihood for a given model LE distribution.] 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. For example, the spread of PVE values across the CCCma LE indicates that even with nearly 70 years of data, there is a roughly 10% uncertainty range on the PVE by the NAO due to the finite record length. This fact is important to consider when comparing the model LE to the observational record, which by definition provides only one value (for a particular observational data set). This one observational value is also subject to the same type of temporal sampling uncertainty as the model (i.e., the observed value could have been different in a "parallel world" under an alternate temporal sequence of internal variability), but without a longer record, this uncertainty is difficult to quantify. In this example, the observed NAO based on the ERA20C_ERAI data set during 1950-2018 (shown in the second column of Fig. 1) has a PVE of 47%, which lies at the upper end of the distribution of values from the CCCma LE.

The 10th - 90th percentile range of PVE values across a LE shows some model dependence, with a lower range for SMHI-KNMI (34-45%) and a higher range for CCCma (42-53%). Note that these two models have different numbers of ensemble members: 16 for SMHI-KNMI and 40 for CCCma. For this reason, it is advantageous to use percentiles rather than absolute (i.e., minimum/maximum) values to define the range of PVE (and other metrics) across each LE for

more robust model inter-comparison; however, the effects of ensemble size differences should still be considered in any model inter-comparison, for example by subsampling the members from the model with the largest ensemble size to match the number of members from the model with the smallest ensemble size. Such exercises can also be performed with the output datasets generated and stored in the CVDP-LE data repository.

The maps in the second column of each row show the NAO regression map based on the first observational PSL data set provided by the user. Like the model LEs, the title and subtitles provide the name of the data set, period of record and PVE. It is clear that all models have a recognizable NAO pattern, with opposite sign between high latitudes and the Atlantic and Pacific midlatitude centers-of-action. However, the relative amplitudes over the Atlantic and Pacific show considerable model dependence: for example, CCCma and NCAR (CESM1) show comparable amplitudes in these two regions, while MPI and CSIRO show larger magnitudes in the North Atlantic compared to the North Pacific. Note, however, that individual ensemble members may differ substantially from their ensemble averages. 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. For example, CCCma shows an average $r = 0.89$, while MPI shows an average $r = 0.97$. The r values for each individual ensemble member may be found in the *Individual Member* view of the CVDP-LE.

The maps in the third column of Fig. 1 show the difference between the ensemble-mean of the model LE and observations (as indicated in the panel title). The maps in the fourth column of Fig. 1 show the percentile rank of the observed value relative to the spread of values across the

members of the model LE. For example, for CCCma (top row), the NAO regression values in the vicinity of the Aleutian Islands exceed observations in every single member of the model ensemble (i.e., the observed rank < 0%: dark blue shading). Conversely, every single ensemble member of the CCCma LE shows NAO regression values in the central North Atlantic that are lower than observations (observed rank > 100%: dark green shading). Observed ranks <10% and >90% highlight a likely bias in the model's simulation of the observed NAO, taking into account sampling fluctuations due to the finite period of analysis. 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.

As a summary metric of the observed percentile rank map, the percentage of the map area with a rank between 10-90% is indicated in the upper right corner (for CCCma, this value is 36%). This summary metric is a useful guide to the overall fidelity of the simulated NAO pattern, taking into account all of the ensemble members. It can be used to quickly compare the overall performance of each model LE's depiction of the NAO. For example, for the set of model LEs in this comparison, GFDL-CM3 has the largest areal percentage of values within 10-90% (67%) and CCCma has the smallest (36%).

The information described in the example above is provided for every spatial pattern shown in the *Ensemble Summary* view of the CVDP-LE, including climatological averages, standard deviations, trends, and modes of variability. Note that for the ENSO spatial composites, the pattern correlations are displayed at the lower right of each model panel. In the case of the SST/TAS/PSL ENSO spatial composites, the pattern correlations are displayed for all 3 variables, separated by a forward slash. For example, the *r* values of the DJF^{+1} El Nino – La Nina composite in the CCCma LE are 0.83/0.46/0.78 for SST/TAS/PSL, respectively (note that TAS is land-only). The number

of El Nino (EN) and La Nina (LN) events that go into each spatial composite (and hovmuller 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).

Interpreting timeseries and derived quantities

Next, we discuss how to make use of the temporal information provided in the "Ensemble Summary" view of the CVDP-LE, including timeseries and derived quantities such as power spectra, wavelets and lag-autocorrelations. We use two examples from the 1950-2018 MMLEA comparison to illustrate the salient information provided in the timeseries displays: annual global-mean surface air temperature anomalies (GMST) and the DJF NAO Index (see Figs. 2 and 3, respectively and the following urls: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/tas_global_avg_ann.summary.png and http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nao_timeseries_djf.summary.png).

The panels in all timeseries displays are arranged in the same order as the spatial maps described above, i.e. alphabetically by model name. The panel titles identify the model and number of ensemble members used in the analysis, with the same color-coding convention as for the spatial maps. 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. The thick gray curve, repeated in each panel, shows the observed timeseries. The observational dataset, period of record and linear trend value are identified at the top of the page in gray font. The trio of values separated by forward slashes in the upper right corner of each panel denotes the 10^{th}, 50^{th} and 90^{th} percentile values of the linear trends across the model LE. The

number in the lower left of each model panel shows the percentage of time that the observed value lies within the 10th-90th percentile range of the model LE values. Note that the y-axis range may differ between the individual panels: this was done to ensure that the model timeseries are clearly visible. However, to facilitate visual comparison across the models, the bottom right panel shows the ensemble-mean timeseries from all the model LEs, color-coded for clarity, along with the observed timeseries in gray.

The information shown in the timeseries display can be used to address the following questions: 1) how does the observed timeseries compare with the externally-forced component (estimated by the ensemble-mean) in any given model LE?; 2) to what extent does the ensemble spread of any given model LE encompass the observed timeseries?; 3) how does the observed trend compare with the forced trend and ensemble spread of trends in any given model LE?; and 4) how do the forced timeseries and distribution of trends compare across different model LEs? The answer to 1) can be used to make a model-informed assessment of the relative contribution of external forcing to the observed timeseries, which contains both unforced (internal) and forced components. The answer to 2) can be used to assess whether the model has a realistic statistical distribution of the combined effects of internal variability and forced climate change in the timeseries. The answer to 3) can be used to make a model-informed assessment of the relative contributions of external forcing and internal variability to the observed trend, and to address the realism of the model's statistical distribution of trends under different permutations of internal variability. Finally, the answer to 4) provides a quantitative comparison of models' forced responses and trends for the timeseries considered.

To be more concrete about how to interpret the information shown in the timeseries displays, we begin by discussing the example of GMST (Fig. 2) followed by the NAO (Fig. 3). Of

all the models, CCCma shows the largest ensemble-mean (i.e., externally-forced) GMST warming trend during 1950-2018 (1.51 C per 69 years), and CSIRO the smallest (0.72 C per 69 years). There is a > 90% likelihood that CCCma's GMST trend is unrealistically large, as seen by the fact that the observed trend value (0.90 C per 69 years) lies well below the 10th percentile of trend values (1.43 C per 69 years) across the individual CCCma members. [The "Individual Member" view of the CVDP-LE can be used to obtain the full range of trend values across the model's ensemble.] Note that conclusions regarding likely model biases in GMST trends (and trends in other quantities) must also take into account whether the model's variability about the trend has a realistic amplitude (in analogy with the previous discussion on the rank of the observed NAO spatial pattern with respect to the model LE ensemble spread). All the other models, including CSIRO, encompass the observed trend value within their 10^{th}-to-90^{th} percentile ensemble spreads. Further, there is a > 90% likelihood that CCCma does not realistically represent the combined effects of external forcing and internal variability on GMST after 2006 since the observations lie outside of the model's distribution. Indeed, the 10^{th}-to-90^{th} percentile ensemble spread in CCCma encompasses the observed GMST value only 45% of the time during 1950-2018, indicating a likely model bias, and cautioning against using this model for quantitative attribution of forced *vs.* internal components of observed GMST change during the past 69 years. Other models perform better in their simulated GMST timeseries. For example, MPI shows that 87% of the time, observations lie within its 10th-to-90th percentile ensemble spread of GMST values, followed closely by GFDL-ESM2M at 83% and NCAR and CSIRO at 80% each. Not surprisingly, the models that perform well in the temporal percentage metric also tend to perform well in their trend simulation.

Having established which models have a credible simulation of the GMST timeseries (i.e., encompassing the observations a high percentage of the time) and GMST trends (i.e., encompassing the observed value within their 10^{th}-to-90^{th} percentile distributions), we can then use them to attribute the observed GMST trends, in particular the partitioning into forced and internal components. Using MPI as an example, one would conclude that most of the observed GMST trend (0.91 C per 69 years) is forced (ensemble mean GMST trend = 0.98 C per 69 years), and that internal variability cooled the earth by 0.07 C during 1950-2018, assuming the forced and internal contributions can be summed up linearly. Quantitatively similar values are found based on GFDL-ESM2M and NCAR. Note that one would not want to use CCCma or GFDL-CM3 for this attribution calculation, since both modes fail to produce a realistic simulation of GMST according to the metrics cited. Indeed, the GMST timeseries from these two models are clearly different from those of the other models and the observations (see the comparison panel in the lower right panel of Fig. 2).

The DJF NAO Index provides a contrasting example to GMST in terms of the relative contributions of forced vs. internal components in the model LEs (Fig. 3 and http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nao_timeseries_djf.summary.png). Unlike GMST, the models exhibit negligible trends in their ensemble-mean (i.e., forced) DJF NAO timeseries (blue curves) compared to the range of trends across each model LE (triad of numbers in the upper right of each panel). For example, CCCma shows a 10^{th}-to-90^{th} percentile range of trends between -0.61 per 69 years and +0.41 per 69 years, with a median value of -0.09 per 69 years (note the PC-based NAO Index is unitless). In other words, the forced component of the DJF NAO trend in CCCma is not significantly different from zero at the 10% confidence level. Indeed, none of the models in the MMLEA show a

significant forced DJF NAO trend during 1950-2018. The observed DJF NAO trend during 1950-2018 (1.11 per 69 years based on ERA20C updated with ERAI) is substantially larger than the 90th percentile trend in any model LE (GFDL-CM3 has the highest 90th percentile trend at 0.72 per 69 years, which is considerably smaller than the observed trend). This suggests that these models are likely deficient in their DJF NAO trend amplitudes during 1950-2018, and thus cannot be used to interpret the relative contributions of forced response *vs*. internal variability in this case. A similar conclusion was reached by Simpson et al. (2018) for a slightly different measure of the NAO in late winter, and more generally, studies have suggested that models may underestimate the low-frequency (decadal and longer) variability of the winter NAO (Kravtsov 2017; Wang et al., 2017). [One could assess the low-frequency variability of the NAO (and other indices) directly in the CVDP-LE by first low-pass filtering the data and then applying the CVDP-LE.]

Scanning the full range of ensemble members (http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nao_timeseries_djf.indmem.png), one finds that there is one ensemble member of one model (SMHI-KNMI) that simulates a realistic DJF NAO trend (1.07 per 69 years), especially if observational uncertainty is taken into account (the trend based on CERA20C is 1.06 per 69 years). Thus, according to the 16-member SMHI-KNMI LE, the probability of simulating a trend of the observed magnitude, based on modeled representation of forced response and internal variability, would be small (one chance in 16, or 6%), but not impossible. Unlike the NAO trend component, most of the models show realistic amplitudes of interannual variability of the NAO, as can be seen by the fact that the observed NAO timeseries (gray curve) resides within the 10th-to-90th percentile envelope (light blue shading) of each model LE a large fraction of the time (ranging from 68% in SMHI-KNMI to 84% in MPI).

Next, we discuss how to interpret quantities derived from the timeseries such as power spectra, wavelets and lag-autocorrelations provided in the "Ensemble Summary" portion of the CVDP-LE. Power spectra are computed for the monthly ENSO, AMV and PDV timeseries based on linearly detrended data. We chose not to include spectra of the atmospheric mode timeseries since they generally follow a "white noise" process and are therefore less useful; however, they are available in the CVDP.

To illustrate the information provided in the "Ensemble Summary" power spectrum plots, we use the example of the Niño3.4 SST Index from the 1950-2018 MMLEA comparison, see: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nino34.powspec.summary.png (Fig. 4). All spectra are shown in variance-preserving form (i.e., linear in both frequency and power). The spectra are computed using the Fast Fourier Transform method and smoothing N adjacent spectral estimates, where N = 7*T*/100 and *T* is the length of the timeseries in years. The panels are arranged in the same way as those in the timeseries display. The x-axis is labeled in units of both frequency (cycles per month; bottom axis) and period (years; top axis) for ease of interpretation. The y-axis is labeled in units of power (degrees C^2 / cycles per month). Note that the y-axis of each panel may differ between models to accommodate their different ENSO amplitudes. However, just as for the timeseries plots, the ensemble-mean spectra from all model LEs are superimposed in the lower right panel for easy comparison. In each panel, the grey curve shows the observed power spectrum (in this case, based on ERSSTv5 1950-2018 as indicated in the plot subtitle), the thick blue curve shows the average of the spectra from each of the model's ensemble members, and the dark and light blue shading show the 10th-to-90th percentile and 25th-to-75th percentile ranges, respectively, across the ensemble. Note that the ensemble-mean of the spectra should not be interpreted in the same way as the ensemble-mean of

the GMST timeseries discussed above. In the timeseries case, the ensemble-mean denotes the forced response. In the power spectrum case, the ensemble mean denotes the average of the spectra of all the individual members, not the spectrum of the ensemble-mean Nino3.4 SST timeseries.

Taking the 40-member CCCma LE (top left panel) as our example, we see that the observed Niño3.4 SST power spectrum lies within the model's ensemble spread at (nearly) all frequencies, indicating that the model simulates realistic amplitude and frequency characteristics of Niño3.4 SST variability. The maximum power in the CCCma Niño3.4 SST index occurs at periods between about 2.5 – 7 years (averaged across the 40 ensemble members; blue curve), similar to observations although with a slight overestimation in the 2.5 – 3 year band. There is considerable spread in the power spectrum across the individual ensemble members, ranging from about 10 degrees C^2 for the 10th percentile to about 35 degrees C^2 for the 90th percentile in the 2.5 – 7 year band. Note that the blue shading only gives the range of power at each frequency band, but does not show whether the shape of the spectrum varies across the individual members (this information can be obtained from the analogous power spectrum plots in the "Individual Members" portion of the CVDP-LE). The other models show varying degrees of fidelity in their Niño3.4 SST power spectra: many overestimate the observed amplitude (most notably, GFDL-ESM2M by a factor of 3 on average). SMHI_KNMI in an outlier in this regard, underestimating the peak ENSO power by approximately a factor of 5.

The Niño3.4 SST power spectrum may be misleading if the model does not correctly simulate the spatial pattern of ENSO. The user can consult the SST standard deviation maps or the ENSO composite maps to determine whether the model has a realistic geographical distribution of interannual SST variability in the tropical Pacific. A case in point is the CSIRO model, which simulates a maximum in DJF (the peak season for ENSO) SST variance over the far western

equatorial Pacific, at odds with the eastern Pacific maximum in observations (http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/sst_spatialstddev_djf.summary.png). This is also evident in the ENSO composite maps for DJF, see: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/nino34.spatialcomp.summary.djf1.png. Thus, the apparently realistic Niño3.4 SST power spectrum in CSIRO belies an erroneous spatial pattern for ENSO. Conversely, SMHI_KNMI exhibits a realistic spatial configuration of SST variance but strongly underestimates its amplitude. The comparison of the ensemble-mean spectra from each model (lower right panel) highlights the wide range of maximum amplitudes, and to a lesser extent, frequency bands associated with dominant spectral peak. In this example, GFDL-ESM2M exhibits the greatest Niño3.4 SST power (about 85 degrees C^2 / cycles per month) and SMHI_KNMI the least (about 5 degrees C^2 / cycles per month), with CCCma showing the most realistic peak values (about degrees 23 C^2 / cycles per month). The maximum power is shifted to slightly higher frequencies than observed in CCCma and GFDL-CM3, while it is shifted to lower frequencies in MPI. Given that these ensemble-mean spectra are based on averaging a large number of members, their statistical characteristics are likely to be robust.

Interpretation of the additional quantities derived from the monthly Niño3.4 SST anomaly timeseries (i.e., standard deviations as a function of month, time-evolving standard deviations based on 30-year running windows, wavelets and lag-autocorrelation curves) follow the same principles as outlined above for the power spectrum.

Interpreting Metrics Tables and Graphics

The CVDP-LE provides a synthesis of model performance based pattern correlations and rms errors for 11 key spatial metrics of climate variability, subjectively but judiciously chosen, as well as an overall benchmark of model performance for all 11 metrics combined. This information is provided in both tabular format and graphical display, as well as output files in ascii and netcdf format. The 11 spatial metrics span a range of oceanic and atmospheric modes of variability, including ENSO, AMV, PDV, atmospheric modes, and standard deviations. Four metrics are used for ENSO, reflecting its overall importance to global climate variability: El Niño minus La Niña global composite maps of TAS and PSL during the peak season (DJF), and composites of equatorial Pacific SST anomalies as a function of time and longitude for El Niño and La Niña separately (in view of their observed asymmetric duration). AMV and PDV are represented by global SST regression maps onto their respective index time series (10-year low-pass filtered AMV' and unfiltered PDV'). Atmospheric modes are represented by NH (north of 20N) PSL regression maps for the NAO (JFM), PNA (DJF) and SAM (DJF). Standard deviations are based on global maps of annual SST and PR.

Here we provide a brief look at the graphical metrics display of pattern correlations, see: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.ensemble.pc.html (Fig. 5). Pattern correlations provide a convenient way to summarize the degree of spatial similarity between two maps, for example the observed and simulated NAO maps discussed earlier. To compute the pattern correlation coefficient (denoted '*r*'), the fields are first normalized by their spatial standard deviation, the areal-mean value is then subtracted from each grid box, and these residual values are then weighted by the square-root of area; then, just as in temporal correlation, the normalized weighted residual values for each field are multiplied together at each grid box, the product is summed over all grid boxes, and this sum

is then divided by the number of grid boxes and the square-root is taken. The resulting *r* value ranges between -1 and 1, where 1 denotes a perfect match between the spatial variations in each field, -1 also denotes a perfect match but the fields have opposite sign, and 0 indicates there is no resemblance between the two fields. In the CVDP-LE, the *r* values for each metric are computed between each individual ensemble member and the first observational data set specified by the user. Each panel shows a different metric, with the models listed along the y-axis and the pattern correlations along the x-axis; note the different x-axis ranges for each metric. Each *r* value is marked with a thin vertical bar, color-coded by model. The "fuzzy" taller vertical bars show the 10th, average and 90th percentile values for each model LE and metric, providing a quick visual summary of how the distributions across the different model LEs compare. The top row of bars (in gray) shows the pattern correlations between the first observational data set and any additional observational data sets specified in the "Namelist" (in this example, 3 additional observational data sets for each variable were specified): their spread provides an indication of observational uncertainty (either to the choice of data set or time period).

The *r* values for a given metric show a wide range across the members of each model LE, attesting to the effect of finite sampling of internal variability in any given simulation. For the example of the NAO, CSIRO shows the widest range of *r* values (0.63-0.90) and SMHI-KNMI the narrowest range (0.90-0.96). These numerical values can also be obtained from the color-coded "Individual Metrics" Table: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.table_0.html and Table 1. Note that these ranges may be affected by the different sample sizes in the two models: 16 for SMHI-KNMI and 30 for CSIRO. For this reason, as mentioned earlier, it is helpful to compare the 10th percentile, 50th and 90th percentile values instead of the absolute ranges. For CSIRO, these values are 0.68, 0.80 and

0.85, respectively; and for SMHI-KNMI they are 0.90, 0.93 and 0.96, respectively (see the numerical values in the color-coded "Ensemble Metrics" Table: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.table_0E.html) and Table 2. The observational uncertainty is small, with all *r* values exceeding 0.96. This highlights that a single ensemble member from a given model is not sufficient to assess the fidelity of the simulated NAO pattern, even when based on nearly 70 years of record. Relatedly, these results also show that the single estimate of the NAO pattern from the 1950-2018 observational record may also be subject to sampling fluctuations (i.e., the "true" NAO pattern, obtained from a hypothetically infinite timeseries, may look slightly different from the pattern estimated from the last 69 years). For further reading on this point, see Deser et al. (2017a) and Deser et al. (2017b) for the NAO and ENSO teleconnection patterns, respectively.

To complement the pattern correlation metrics, the CVDP-LE also provides *rms* errors (*rmse*) for the simulated patterns with respect to observations, in both graphical form (http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.ensemble.rms.html) and tabular form (Ensemble Summary Table: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.table_14E.html; and Individual Metrics Table: http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE_repository/MMLEA_1950-2018/metrics.table_14.html and Tables 3 and 4). The calculation of *rmse* is similar to that for *r* values, except that the values are not normalized and the difference between the model and observed value at each grid point is used in place of the product. The *rmse* thus provides a summary of the relative amplitudes of the observed and simulated patterns. Note the smaller the rmse, the "better" the model. For the NAO example, the upper end of the range of *rmse* values is greatest for CCCma and CSIRO. Thus, it is clear that a lower *r* value

does not necessarily correspond to a higher *rmse*, since the *r* values are generally substantially higher for CCCma than CSIRO. This highlights the importance of considering both pattern correlation and rms error in model performance.

The CVDP-LE provides an overall benchmark of model performance or "Mean Score", computed by taking the average of the 11 *r* values (after applying a Fischer z-transform) and the 11 *rmse* values (after normalizing each by the spatial rms of the observed pattern to account for the different units of each variable). The results, shown in the lower right panel of the metrics graphical display and in the right-hand column of the metrics Tables, provide a quick summary of model performance and model inter-comparison. For the 1950-2018 MMLEA, the Mean Score *r* value provides an effective way to discriminate among the 7 model LEs: SMHI-KNMI and CSIRO clearly fall at the lower range of model performance (10^{th}-to-90^{th} percentile ranges of 0.62 - 0.70 and 0.70 - 0.75, respectively) while the remaining 5 models are all clustered in the range 0.77 – 0.84). The Mean Score *rmse* shows somewhat less inter-model variation, and the models that score well in their *r* value do not necessarily score well in *rmse* and vice versa (compare Tables 2 and 4). We emphasize that this "Mean Score" of model performance is heavily weighted toward ENSO, and that there may be interdependencies amongst the 11 metrics used.

A novel feature of the metrics Tables is the ability to sort the values in a particular column by selecting the appropriate link above the Table. For example, if the user wants to rank the models according to their simulation of the NAO pattern, they would select the "NAO" link. The NAO column would be outlined by a thick black rectangle to indicate that the ranking is performed for that quantity. The models can also be sorted alphabetically and by Mean Score; the default arrangement ("Namelist (default)") is the order in which the models are listed in the input files. For the MMLEA example, sorting the Pattern Correlation Ensemble Metrics Table by Mean Score

ranks the 90th percentile of GFDL-CM3 at the top of the list (Mean Score *r* value of 0.84) and the 10th percentile of SMHI_KNMI at the bottom (Mean Score *r* value of 0.63). Considering only the average values of the Mean Scores for each model, GFDL-CM3 is still ranked at the top (*r* value of 0.83) and SMHI_KNMI at the bottom (*r* value of 0.67). Sorting the *rmse* Ensemble Metrics Table by Mean Score ranks the 10th percentile of SMHI_KNMI at the top of the list (value of 0.59) and the 90th percentile of GFDL-ESM2M at the bottom (value of 0.94).

5d. "Individual Members" View

The "Individual Members" view provides access to the full set of CVDP-LE diagnostics and metrics for all the individual members of each model LE as well as for each observational data set. Like the "Ensemble Summary" view, this information is displayed graphically on a single page, providing a comprehensive view of member-to-member variations within a given model and between models, as well as between the different observational data sets and/or time periods. The information is also provided in netcdf format for later use. The spatial patterns for the individual ensemble members may be viewed "as is" ("Indiv" link) or as differences from the first observational data set specified by the user ("Bias" link). Metrics Tables are also provided for the individual ensemble members and observational data sets, and can be sorted according to model ("Namelist (default)"), metric and Mean Score.

6. Best practices and tips for applying the CVDP-LE

It is up to the user to decide what set of model LE simulations and observational data sets to input to the CVDP-LE. The user may be interested in comparing a particular set of model LEs and observational data sets for the same time period (as in the MMLEA example above), or they

may wish to specify different time periods for a particular model LE to assess changes in modes of variability over time, or more generally, the sensitivity of the results to the period of record analyzed. Alternatively, the user may wish to compare different subsets of ensemble members from a particular model LE or multiple LEs, or to compare different temporally-filtered versions of the same data sets to investigate sensitivities to time scale. The user may also decide to apply the CVDP-LE to multiple control simulations from different models, or to different versions or physical hierarchies of the same model. Another application may be to divide a long control simulation into segments to create an "ensemble" of shorter analysis periods.

If the purpose is to isolate the internally-generated component of variability and trends, the ensemble-mean of the model LE can be subtracted from each individual member at each time step and grid box before applying the package. The user may also be interested in examining possible forced changes in internal variability (amplitude and spatial pattern), in which case present-day and future time periods from the same collection of model LEs can be combined in a single CVDP-LE application. The user may also decide to omit observational data for comparison, and instead use a particular model (either a single simulation, LE ensemble-mean or multi-model ensemble-mean) as the reference against which other models are compared.

Alternatively, to test the accuracy of a particular methodology that might be applied to the observational record for "detection and attribution" purposes, for example optimal fingerprinting (Hegerl et al., 1997), dynamical adjustment (Guo et al., 2018) or pattern recognition (Sippel et al., 2019; Wills et al., 2020), the user could use specify each of the LE member in turn as a different "observational" data set and compare their results to the true forced response (given by the LE ensemble mean). The ideas listed above are just some of the ways that the CVDP-LE might be utilized; users will undoubtedly come up with their own creative applications.

In all CVDP-LE model evaluations, the user needs to be aware of observational uncertainties related to instrument error, measurement techniques and inhomogeneities, spatial and temporal coverage, treatment of missing data, and possible application of optimal interpolation or data assimilation methods. In particular, possible mismatches in spatial and/or temporal coverage between observations and models must be considered for proper comparison and interpretation. A useful resource on observational data sets is the Climate Data Guide, which provides expert-user guidance and information on the strengths, limitations and applicability of nearly 200 data sets. The user should also take into account whether the analysis period is long enough to provide adequate statistics on the various modes of variability. For example, longer records will be needed to robustly assess modes of decadal variability compared to modes of interannual variability. Similarly, if the user specifies different time periods for different observational data sets and/or models, direct comparison of some metrics may no longer be meaningful, for example trends. Also, when future time periods from model simulations are analyzed, comparison to observations may in some cases be problematic to interpret.

For added utility, we have created a set of CVDP-LE comparisons and accompanying data repositories accessible from the CVDP-LE webpage (http://www.cesm.ucar.edu/working_groups/CVC/cvdp-le/data-repository.html). These include the MMLEA, and CMIP6 models with at least 10 ensemble members, for three time periods: 1950-2018, 2019-2100, and 1950-2100. We welcome user feedback on any aspect of the CVDP-LE via the CVDP-LE support page (http://www.cesm.ucar.edu/working_groups/CVC/cvdp-le/support.html), including suggestions for additional metrics and diagnostics to include in the package as well as for other comparisons to add to the data repository. We also welcome your comments on this document.

7. Acknowledgements

We thank the members of NCAR's Climate Analysis Section for providing important feedback on the design of the CVDP-LE and for their helpful comments and suggestions on this document. We also appreciate feedback from Claude Frankignoul, Karen McKinnon and Keith Rogers on this document.

8. References

Amaya, D. 2020: Indices of the North and South Pacific Meridional Modes based on sea surface temperature. *In preparation*.

Barnes, E. A., Hurrell, J. W. & Uphoff, I. E., 2019: Viewing forced climate patterns through an AI lens. *Geophys. Res. Lett*. 46, 13389–13398.

Bonfils, C., B. D. Santer, J. C. Fyfe, K. Marvel, T.J. Phillips, and S. Zimmerman, 2019: Human influence on joint changes in temperature, rainfall and continental aridity. *Nature Climate Change*, 2019.

Chiang J.C. and D.J. Vimont, 2004: Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. *J. Climate* 17:4143–58.

Danabasoglu, G., S. G. Yeager, Y. -O. Kwon, J. J. Tribbia, A. S. Phillips, and J. W. Hurrell, 2012. Variability of the Atlantic Meridional Overturning Circulation in CCSM4. *J. Climate*, 25, 5153-5172, doi: 10.1175/JCLI-D-11-00463.1.

Deser, C., M. A. Alexander, S. -P. Xie, and A. S. Phillips, 2010: Sea surface temperature variability: patterns and mechanisms. *Ann. Rev. Mar. Sci.*, **2010.2**, 115-143, doi:10.1146/annurev-marine-120408-151453.

Deser, C., A. S. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527-546, DOI 10.1007/s00382-010-0977-x.

Deser, C., L. Terray and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237-2258, doi: 10.1175/JCLI-D-15-0304.1.

Deser, C., J. W. Hurrell and A. S. Phillips, 2017a: The Role of the North Atlantic Oscillation in European Climate Projections. *Clim. Dyn.*, doi: 10.1007/s00382-016-3502-z.

Deser, C., I. R. Simpson, K. A. McKinnon and A. S. Phillips, 2017b: The Northern Hemisphere extra-tropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *J. Climate*, **30**, 5059-5082, doi: 10.1175/JCLI-D-16-0844.1.

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

Doi, T., T. Tozuka and T. Yamagata, 2009: Interannual variability of the Guinea Dome and its possible link with the Atlantic Meridional Mode. *Climate Dynamics*, 33, 985-998, doi:10.1007/s00382-009-0574-z.

Enfield, D.B., A.M. Mestas, D.A. Mayer, and L. Cid-Serrano, 1999: How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures?, *JGR-O*ceans, 104, 7841-7848.

Eyring, V. et al., 2016: ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP, Geosci. Model Dev.,9, 1747-1802, doi:10.5194/gmd-9-1747-2016.

Fasullo, J., A. S. Phillips and C. Deser, 2020: Evaluation of Leading Modes of Climate Variability in the CMIP Archives. *J. Climate*, in press.

Gould, S. J., 1989: Wonderful Life: The Burgess Shale and the Nature of History (W. W. Norton & Co., 1989).

Guo, R., C. Deser, L. Terray and F. Lehner, 2019: Human influence on winter precipitation trends (1921-2015) over North America and Eurasia revealed by dynamical adjustment. *Geophys. Res. Lett.*, **46**, doi: 10.1029/2018GL081316.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, 1996: Detecting greenhouse-gasinduced climate change with an optimal fingerprint method. *J. Climate*, 9, 2281–2306, https://doi.org/10.1175/1520-0442 (1996)009,2281:DGGICC.2.0.CO;2.

Hurrell, J. W., and C. Deser, 2009: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **78**, No. 1, 28-41, doi:10.1016/j.jmarsys.2008.11.026.

Kay, J. E., C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, M. Holland, P. Kushner, J. -F. Lamarque, D. Lawrence, K. Lindsay, A. Middleton, E. Munoz, R. Neale, K. Oleson, L. Polvani, and M. Vertenstein, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Met. Soc.*, **96**, 1333–1349, doi: 10.1175/BAMS-D-13-00255.1.

Kravtsov, S., 2017: Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. Geophys. Res. Lett., 44, 5749–5757, https:// doi.org/10.1002/2017GL074016

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. Fischer, L. Brunner, R. Knutti and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple Large Ensembles and CMIP5/6. *Earth Sys. Dyn. Discuss., Special Issue on Large Ensembles*, **11**, 491–508, doi: 10.5194/esd-11-491-2020.

Maher, N. et al., 2019: The Max Planck Institute Grand Ensemble – enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst*. 11, 2050–2069.

Mantua. N.J., Hare, S.R., Zhang, Y., Wallace, J.M. & Francis, R.C. (1997). A Pacific interdecadal oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc*., 78, 1069-1079.

Milinski, S., Maher, N., and Olonscheck, D. (2019). How large does a large ensemble need to be? *Earth Syst. Dynam. Discuss.* doi:10.5194/esd-2019-70.

Mo K. C. and R. W. Higgins, 1998: The Pacific–South American modes and tropical convection during the Southern Hemisphere winter. *Mon. Wea. Rev*., 126, 1581–1596.

Phillips, A. S., C. Deser, and J. Fasullo, 2014: A new tool for evaluating modes of variability in climate models. *EOS*, **95**, 453-455, doi: 10.1002/2014EO490002.

Rasmusson E.M. and Carpenter T.H., 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Nino. *Mon. Weather Rev*. 110:354–84.

Saji N.H., Goswami B.N., Vinayachandran P.N., Yamagata T., 1999: A dipole mode in the tropical Indian Ocean, *Nature*, 401, 360-363.

Santer, B., J. C. Fyfe, Susan Solomon, Jeffrey F. Painter, Céline Bonfils, Giuliana Pallotta, Mark D. Zelinka, 2019: Proceedings of the National Academy of Sciences , 116 (40) 19821-19827; DOI: 10.1073/pnas.1904586116

Simpson, I. R., C. Deser, K. A. McKinnon, and E. A. Barnes, 2018: Modelled and observed multidecadal variability in the North Atlantic jet stream and its connection to Sea Surface Temperatures. *J. Climate*, 8313-8338, doi: 10.1175/JCLI-D-18-0168.1.

Sippel, S. et al., 2019: Uncovering the forced climate response from a single ensemble member using statistical learning. *J. Climate* https://doi.org/10.1175/JCLI-D-18-0405.1.

Tél, T., Bódai, T., Drótos, G. *et al.* , 2020: The Theory of Parallel Climate Realizations. *J Stat Phys* **179,** 1496–1530. https://doi.org/10.1007/s10955-019-02445-7

Thompson, D.W.J. & Wallace, J.M., 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*. 13. 1000-1016. https://doi.org/10.1175/1520-0442(2000)01360;1000:amitec62;2.0.co;2.

Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *J. Climate*, **28**, 6443-6456, doi: 10.1175/JCLI-D-14-00830.1.

Torrence, C., and G. P. Compo, 1998: A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61–78, https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.

Trenberth, K. T. and J. W. Hurrell, 1994: Decadal atmospheric–ocean variations in the Pacific. Climate Dyn., 9, 303–319.

Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation Revisited: Sea Level Pressures, Surface Temperatures, and Precipitation. *J. Climate*, **13**, 4358–4365, https://doi.org/10.1175/1520-0442(2000)013<4358:TSORSL>2.0.CO;2.

Trenberth, K.E. and Shea, D.J., 2006. Atlantic hurricanes and natural variability in 2005. *Geophysical research letters*, *33*(12).

Wang, X., J. Li, C. Sun, and T. Liu, 2017: NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. J. Geophys. Res. Atmos., 122, 4202–4227, https://doi.org/10.1002/2016JD025979.

Wills, R. C. J., D. S. Battisti, K. C. Armour, T. Schneider, and C. Deser, 2020: Pattern Recognition Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations. *J. Climate*, in press.

Zebiak, S. E., 1993: Air–sea interaction in the equatorial Atlantic region. *J. Climate*, 6, 1567–1586.

Zhang, H., A. Clement, and P. DiNezio, 2014: The South Pacific meridional mode: A mechanism for ENSO-like variability, *J. Clim*., 27, 769–783, doi:10.1175/JCLI-D-13-00082.1.

9. Figures and Tables

All figures and tables are taken from http://webext.cgd.ucar.edu/Multi-Case/CVDP-LE/MMLEA_1950-2018/ .

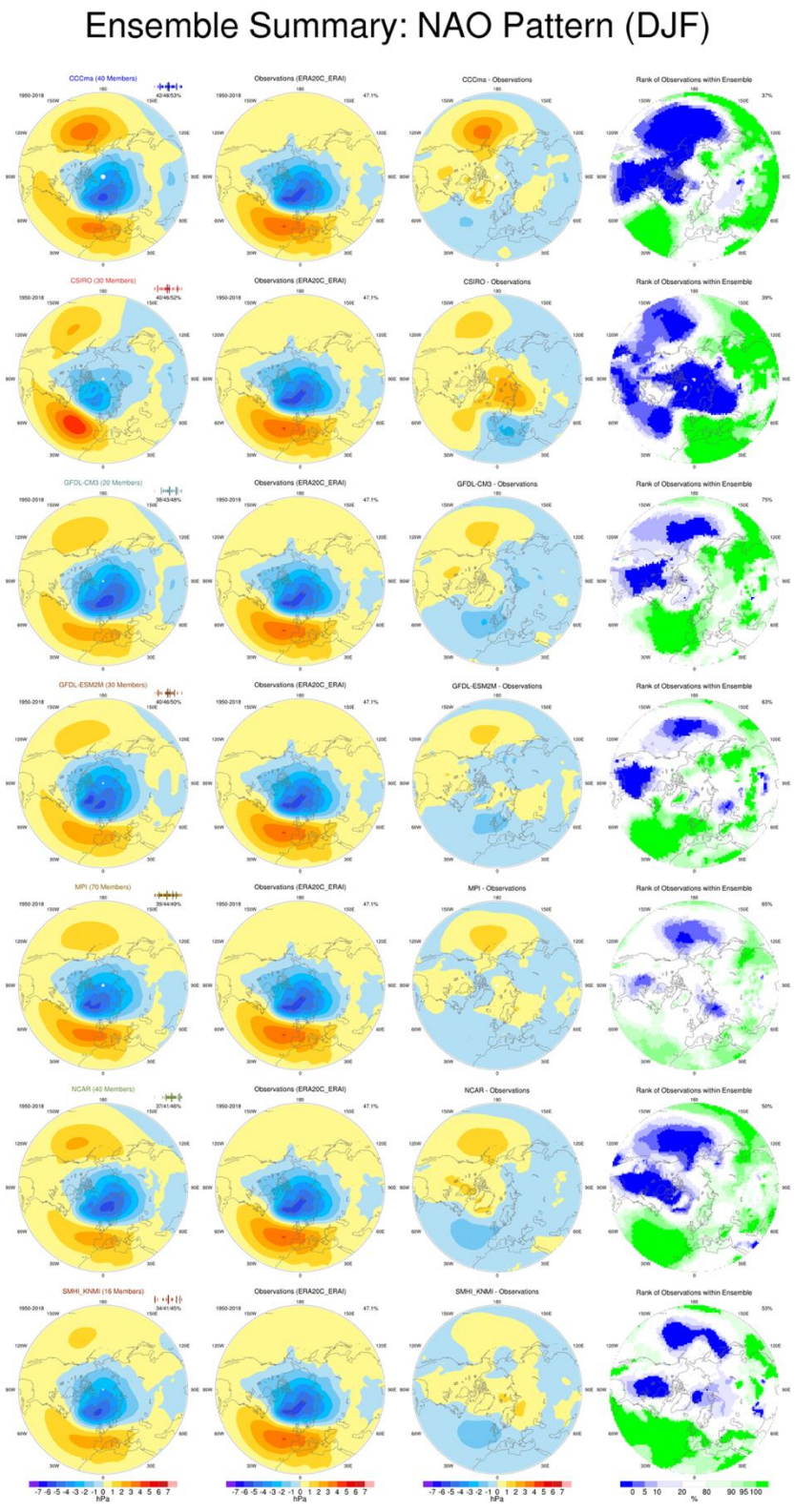
Ensemble Summary: NAO Pattern (DJF)
CCCma (40 Members)
Observations (ERA20C_ERAI)
CCCma - Observations
Rank of Observations within Ensemble
CSIRO (30 Members)
CSIRO - Observations
GFDL-CM3 (20 Members)
GFDL-CM3 - Observations
GFDL-ESM2M (30 Members)
GFDL-ESM2M - Observations
MPI (70 Members)
MPI - Observations
NCAR (40 Members)
NCAR - Observations
SMHI_KNMI (16 Members)
SMHI_KNMI - Observations
hPa
%

Figure 1.

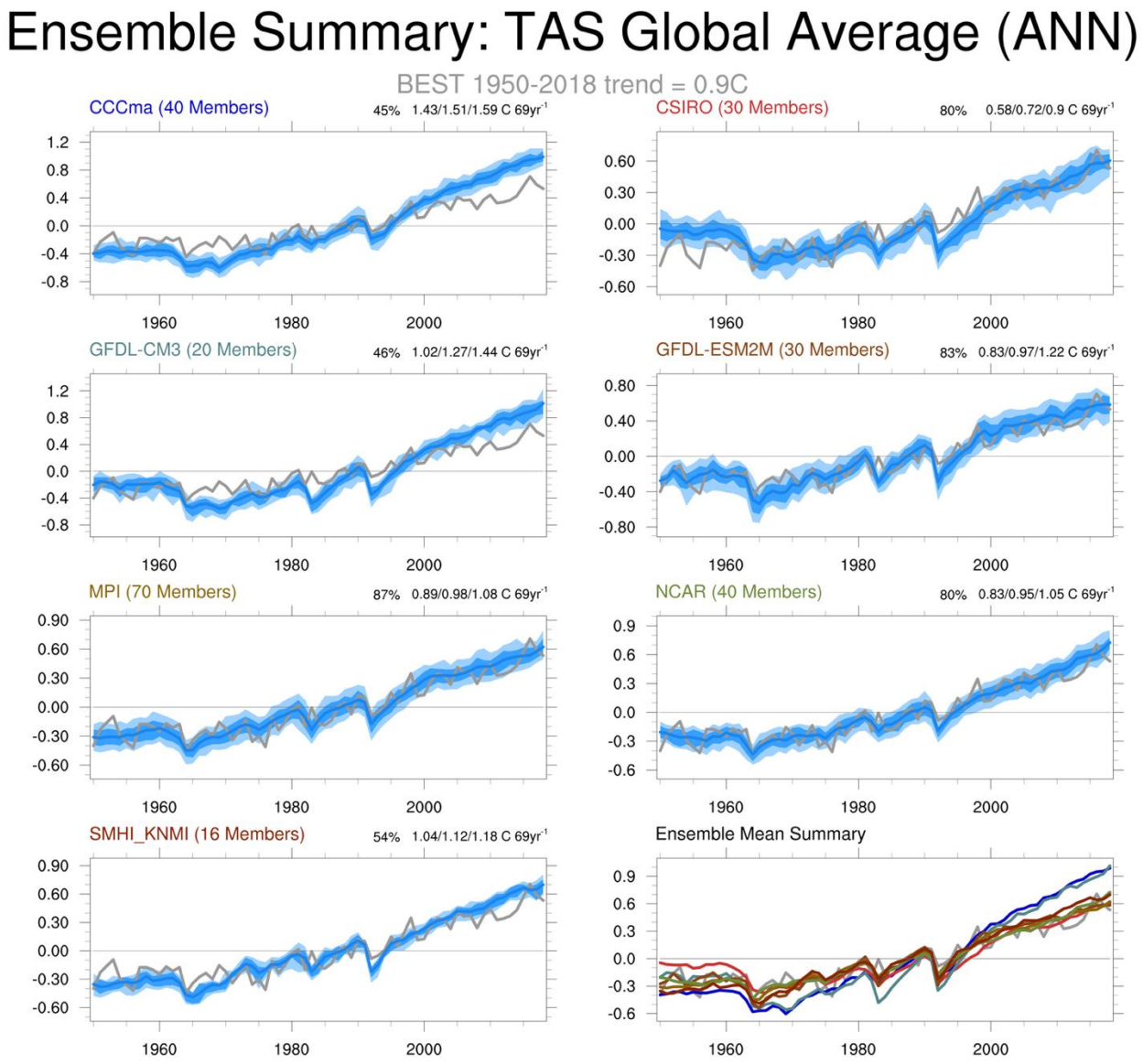
Ensemble Summary: TAS Global Average (ANN)
BEST 1950-2018 trend = 0.9C
CCCma (40 Members)
45% 1.43/1.51/1.59 C 69yr-1
CSIRO (30 Members)
80% 0.58/0.72/0.9 C 69yr-1
GFDL-CM3 (20 Members)
46% 1.02/1.27/1.44 C 69yr-1
GFDL-ESM2M (30 Members)
83% 0.83/0.97/1.22 C 69yr-1
MPI (70 Members)
87% 0.89/0.98/1.08 C 69yr-1
NCAR (40 Members)
80% 0.83/0.95/1.05 C 69yr-1
SMHI_KNMI (16 Members)
54% 1.04/1.12/1.18 C 69yr-1
Ensemble Mean Summary
1960
1980
2000

Figure 2.

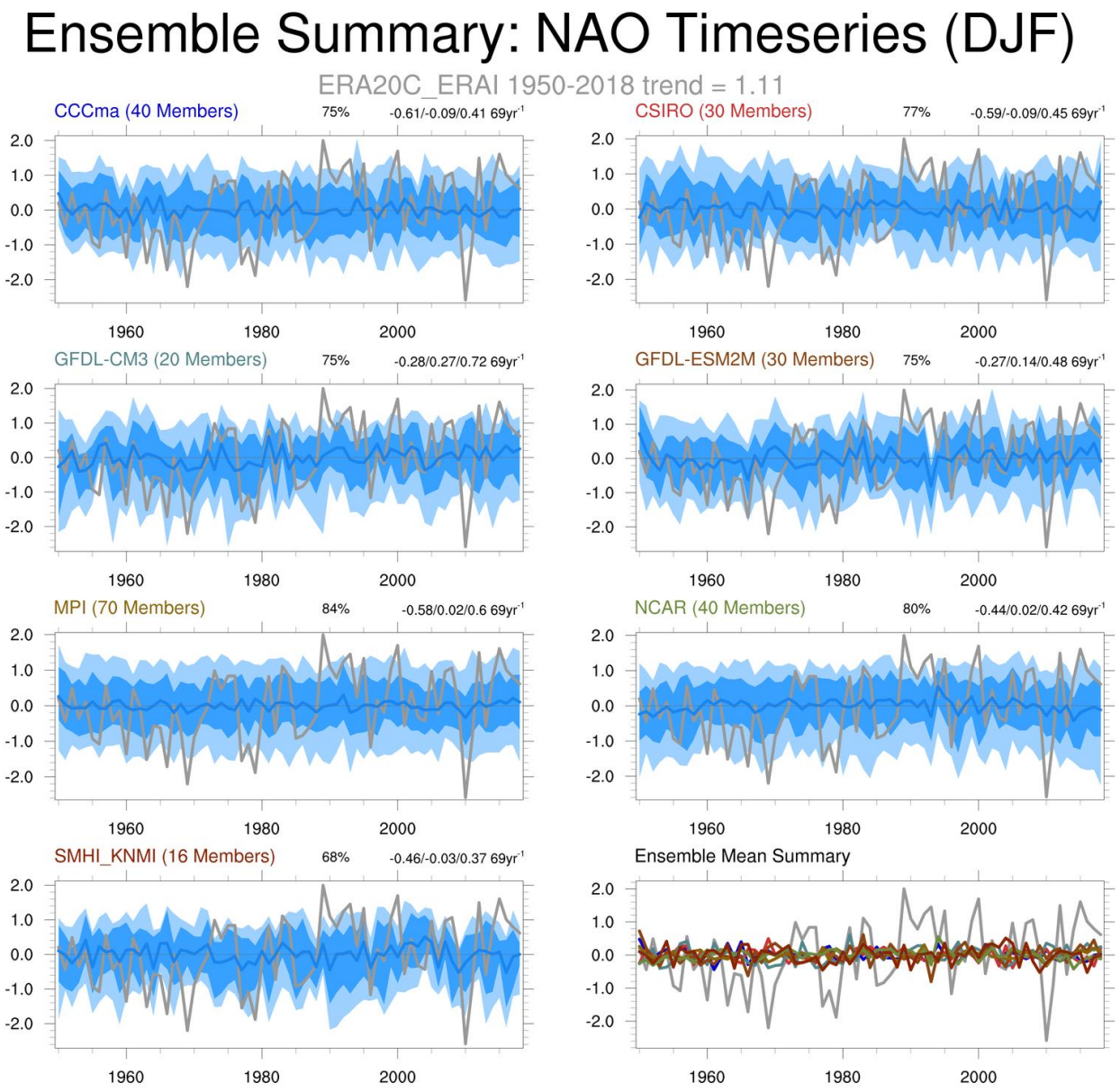
Ensemble Summary: NAO Timeseries (DJF)
ERA20C_ERAI 1950-2018 trend = 1.11
CCCma (40 Members)
75%
-0.61/-0.09/0.41 69yr-1
CSIRO (30 Members)
77%
-0.59/-0.09/0.45 69yr-1
GFDL-CM3 (20 Members)
75%
-0.28/0.27/0.72 69yr-1
GFDL-ESM2M (30 Members)
75%
-0.27/0.14/0.48 69yr-1
MPI (70 Members)
84%
-0.58/0.02/0.6 69yr-1
NCAR (40 Members)
80%
-0.44/0.02/0.42 69yr-1
SMHI_KNMI (16 Members)
68%
-0.46/-0.03/0.37 69yr-1
Ensemble Mean Summary
2.0
1.0
0.0
-1.0
-2.0
1960
1980
2000

Figure 3.

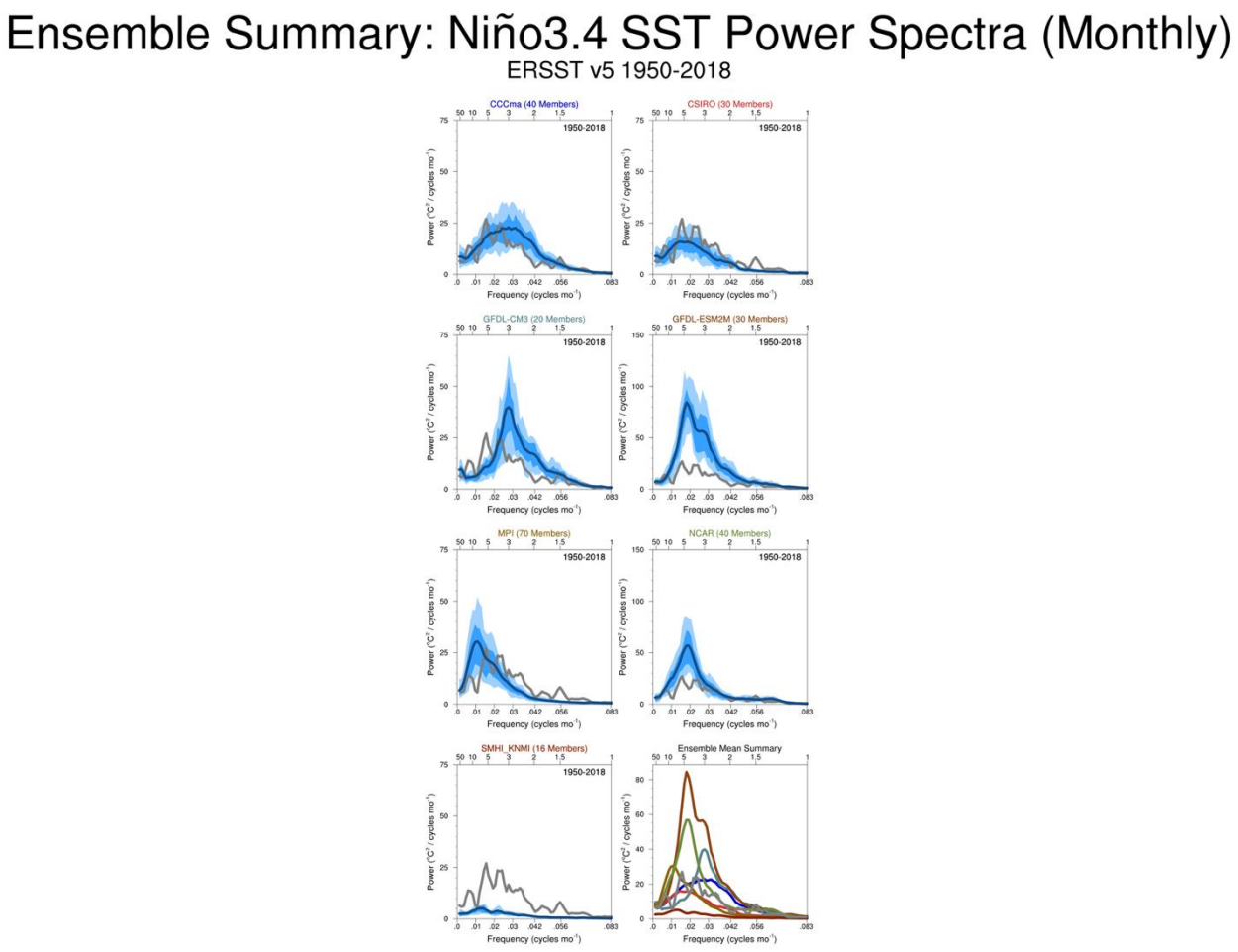
Ensemble Summary: Niño3.4 SST Power Spectra (Monthly)
ERSST v5 1950-2018
CCCma (40 Members)
CSIRO (30 Members)
GFDL-CM3 (20 Members)
GFDL-ESM2M (30 Members)
MPI (70 Members)
NCAR (40 Members)
SMHI_KNMI (16 Members)
Ensemble Mean Summary
1950-2018
Power (°C² / cycles mo⁻¹)
Frequency (cycles mo⁻¹)

Figure 4.

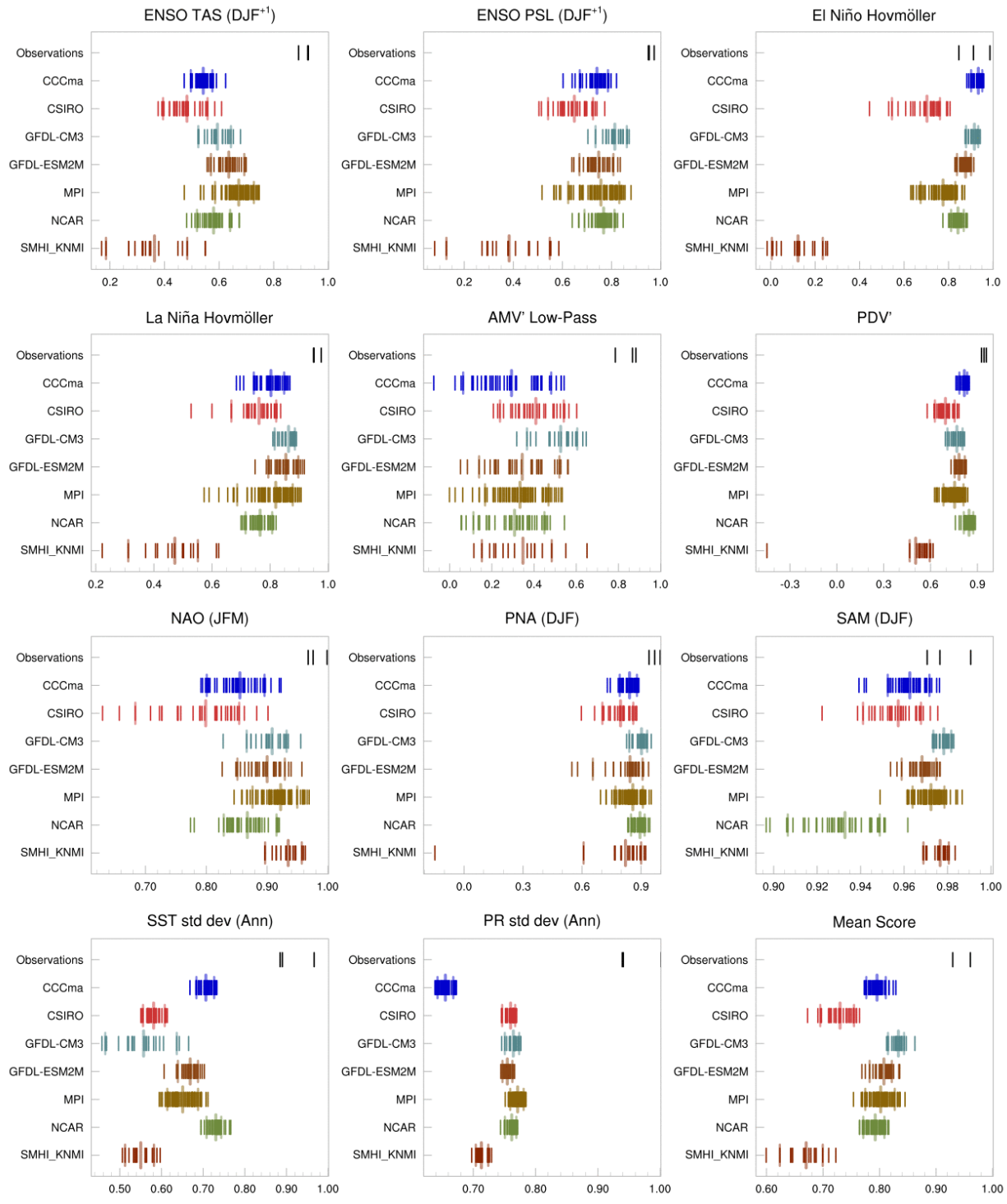
ENSO TAS (DJF+1)
ENSO PSL (DJF+1)
El Niño Hovmöller
La Niña Hovmöller
AMV' Low-Pass
PDV'
NAO (JFM)
PNA (DJF)
SAM (DJF)
SST std dev (Ann)
PR std dev (Ann)
Mean Score
Observations
CCCma
CSIRO
GFDL-CM3
GFDL-ESM2M
MPI
NCAR
SMHI_KNMI

Figure 5.

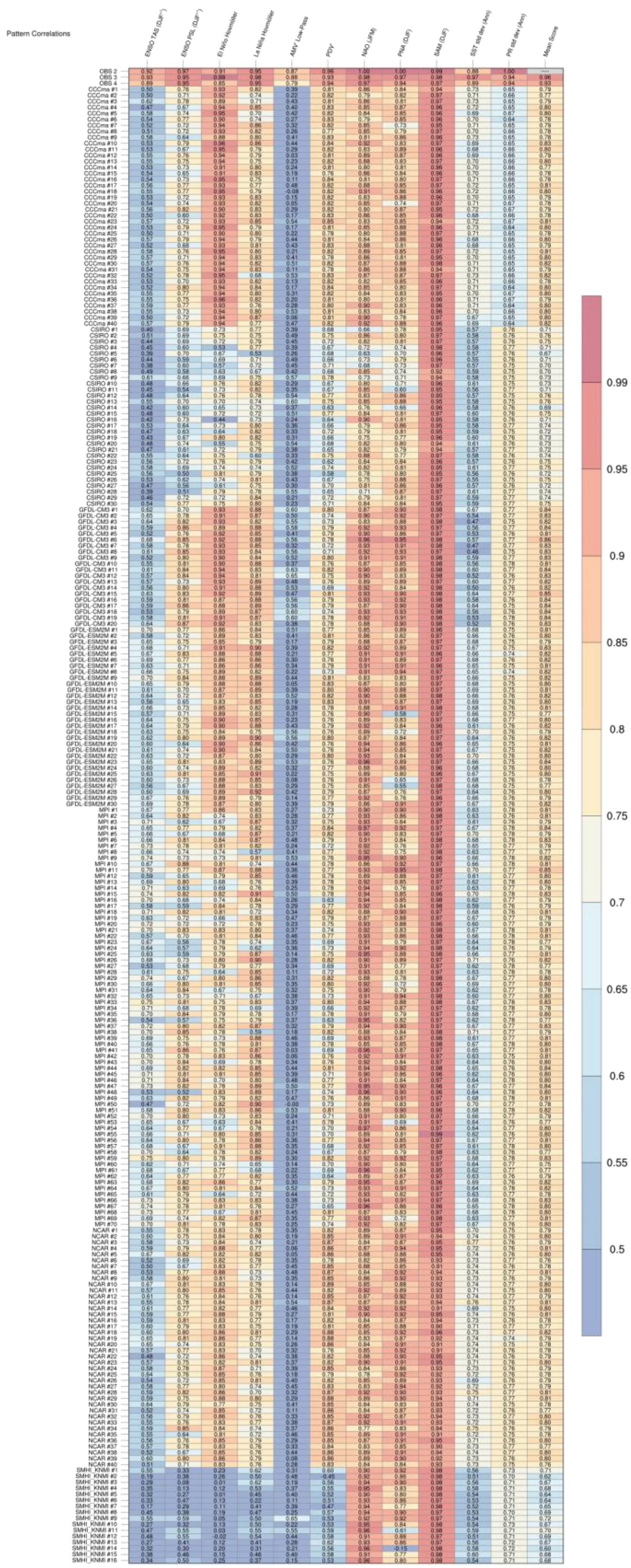
Pattern Correlations
ENSO TAS (DJF⁺)
ENSO PSL (DJF⁺)
El Niño Hovmöller
La Niña Hovmöller
AMV Low-Pass
PDV
NAO (JFM)
PNA (DJF)
SAM (DJF)
SST std dev (Ann)
PR std dev (Ann)
Mean Score
0.99
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5

Table 1.

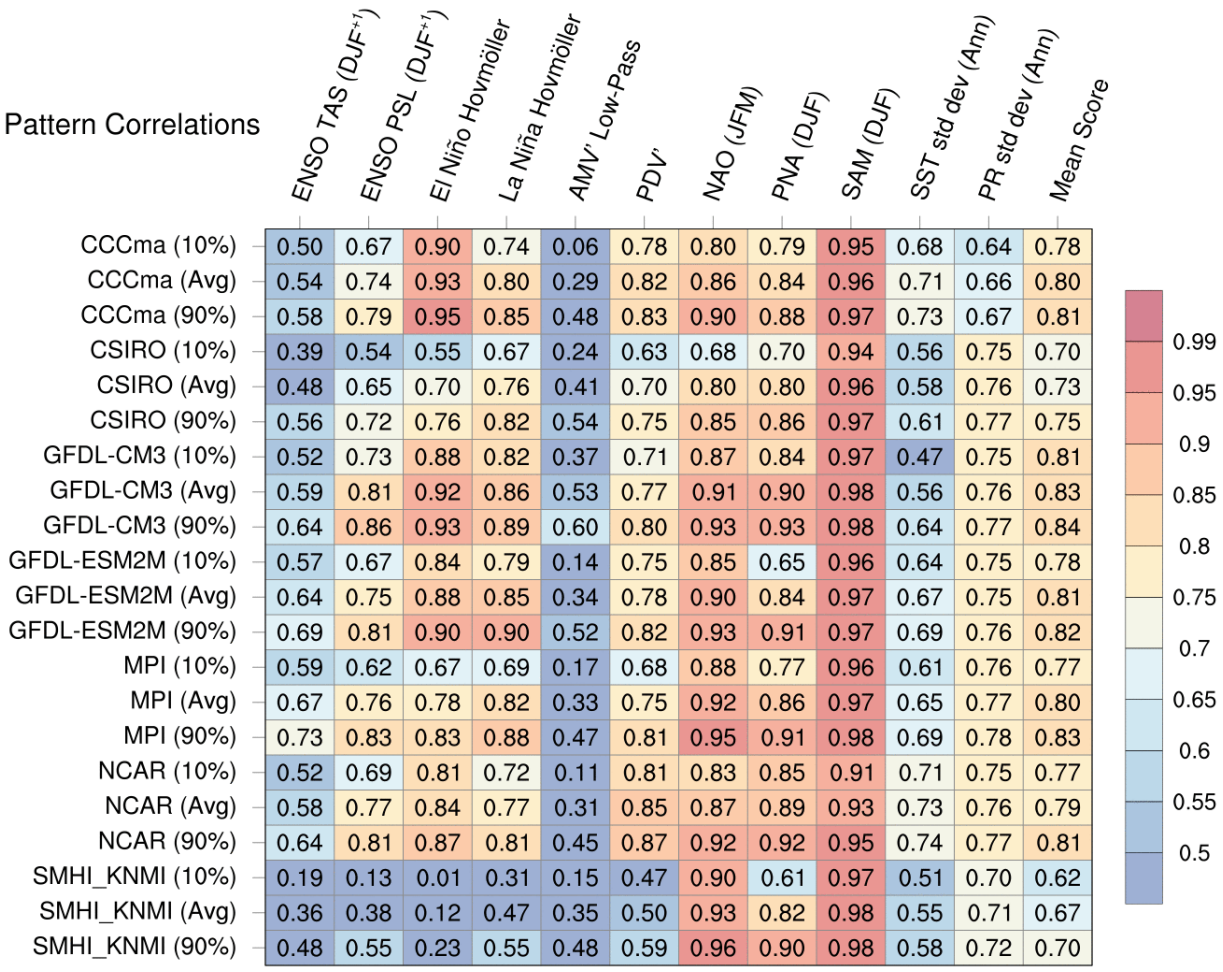
Pattern Correlations
ENSO TAS (DJF+1)
ENSO PSL (DJF+1)
El Niño Hovmöller
La Niña Hovmöller
AMV Low-Pass
PDV
NAO (JFM)
PNA (DJF)
SAM (DJF)
SST std dev (Ann)
PR std dev (Ann)
Mean Score
CCCma (10%) 0.50 0.67 0.90 0.74 0.06 0.78 0.80 0.79 0.95 0.68 0.64 0.78
CCCma (Avg) 0.54 0.74 0.93 0.80 0.29 0.82 0.86 0.84 0.96 0.71 0.66 0.80
CCCma (90%) 0.58 0.79 0.95 0.85 0.48 0.83 0.90 0.88 0.97 0.73 0.67 0.81
CSIRO (10%) 0.39 0.54 0.55 0.67 0.24 0.63 0.68 0.70 0.94 0.56 0.75 0.70
CSIRO (Avg) 0.48 0.65 0.70 0.76 0.41 0.70 0.80 0.80 0.96 0.58 0.76 0.73
CSIRO (90%) 0.56 0.72 0.76 0.82 0.54 0.75 0.85 0.86 0.97 0.61 0.77 0.75
GFDL-CM3 (10%) 0.52 0.73 0.88 0.82 0.37 0.71 0.87 0.84 0.97 0.47 0.75 0.81
GFDL-CM3 (Avg) 0.59 0.81 0.92 0.86 0.53 0.77 0.91 0.90 0.98 0.56 0.76 0.83
GFDL-CM3 (90%) 0.64 0.86 0.93 0.89 0.60 0.80 0.93 0.93 0.98 0.64 0.77 0.84
GFDL-ESM2M (10%) 0.57 0.67 0.84 0.79 0.14 0.75 0.85 0.65 0.96 0.64 0.75 0.78
GFDL-ESM2M (Avg) 0.64 0.75 0.88 0.85 0.34 0.78 0.90 0.84 0.97 0.67 0.75 0.81
GFDL-ESM2M (90%) 0.69 0.81 0.90 0.90 0.52 0.82 0.93 0.91 0.97 0.69 0.76 0.82
MPI (10%) 0.59 0.62 0.67 0.69 0.17 0.68 0.88 0.77 0.96 0.61 0.76 0.77
MPI (Avg) 0.67 0.76 0.78 0.82 0.33 0.75 0.92 0.86 0.97 0.65 0.77 0.80
MPI (90%) 0.73 0.83 0.83 0.88 0.47 0.81 0.95 0.91 0.98 0.69 0.78 0.83
NCAR (10%) 0.52 0.69 0.81 0.72 0.11 0.81 0.83 0.85 0.91 0.71 0.75 0.77
NCAR (Avg) 0.58 0.77 0.84 0.77 0.31 0.85 0.87 0.89 0.93 0.73 0.76 0.79
NCAR (90%) 0.64 0.81 0.87 0.81 0.45 0.87 0.92 0.92 0.95 0.74 0.77 0.81
SMHI_KNMI (10%) 0.19 0.13 0.01 0.31 0.15 0.47 0.90 0.61 0.97 0.51 0.70 0.62
SMHI_KNMI (Avg) 0.36 0.38 0.12 0.47 0.35 0.50 0.93 0.82 0.98 0.55 0.71 0.67
SMHI_KNMI (90%) 0.48 0.55 0.23 0.55 0.48 0.59 0.96 0.90 0.98 0.58 0.72 0.70
0.99
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5

Table 2.

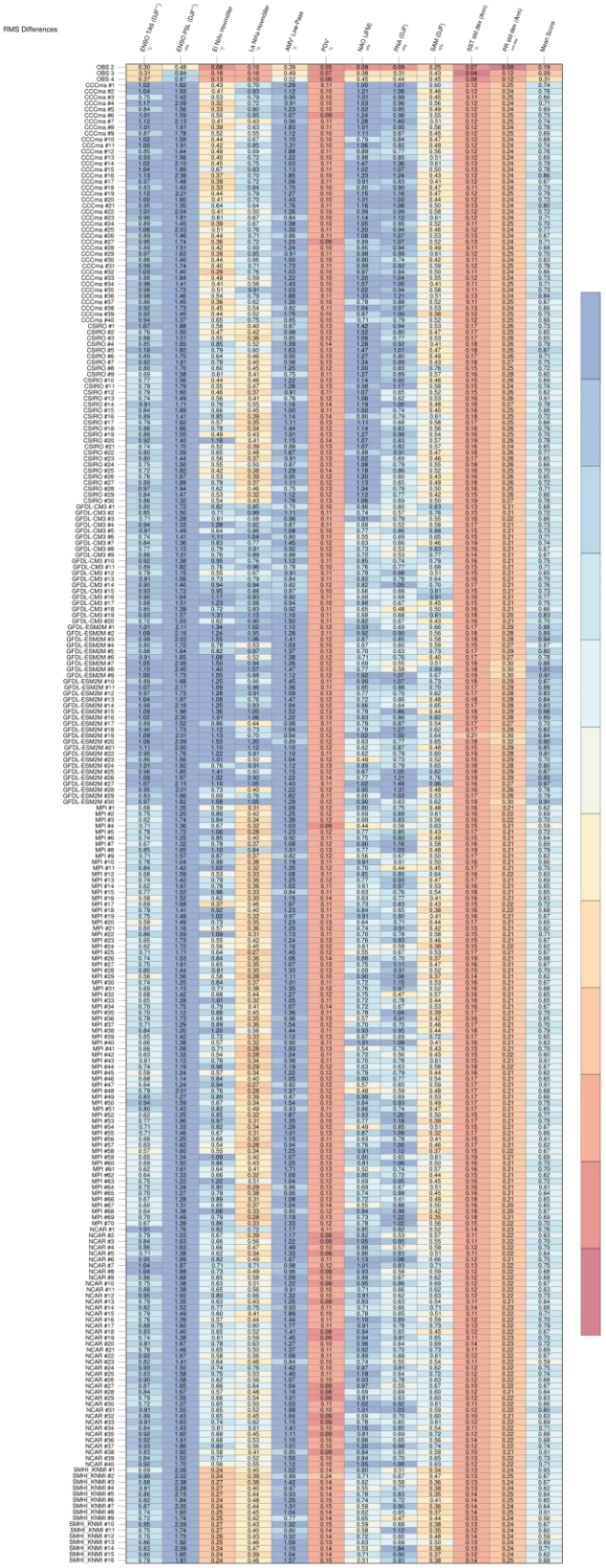
RMS Differences
ENSO 1 std (DJF)
ENSO PSL (DJF)
El Niño Anomalies
LA Niña Anomalies
AMV, Low Pass
PDV
NAO (JFM)
PNA (DJF)
SAM (DJF)
SST, std dev (Ann)
PR std dev (Ann)
Mean Score
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.05

Table 3.

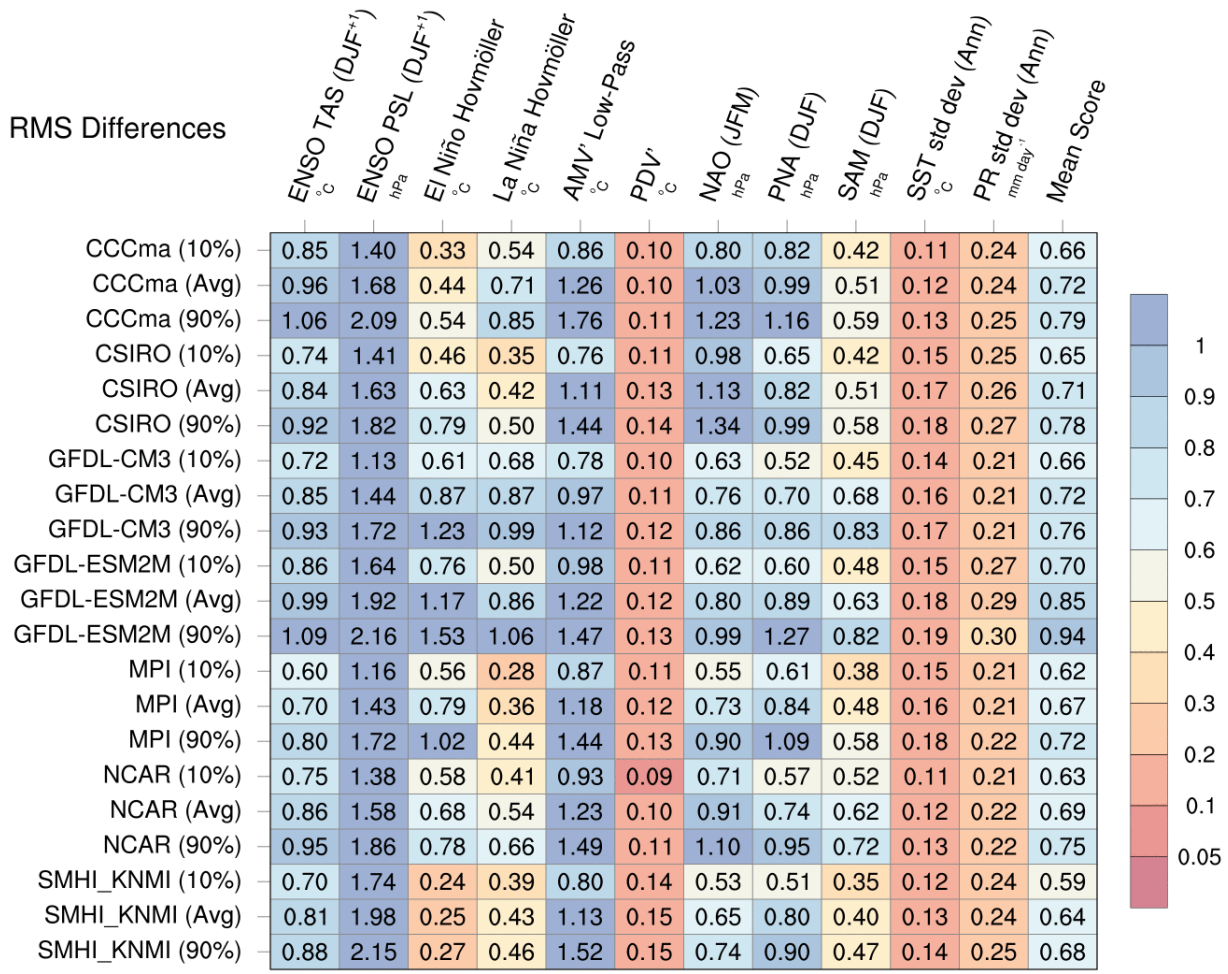
RMS Differences
ENSO TAS (DJF+1) °C
ENSO PSL (DJF+1) hPa
El Niño Hovmöller °C
La Niña Hovmöller °C
AMV' Low-Pass °C
PDV' °C
NAO (JFM) hPa
PNA (DJF) hPa
SAM (DJF) hPa
SST std dev (Ann) °C
PR std dev (Ann) mm day-1
Mean Score
CCCma (10%) 0.85 1.40 0.33 0.54 0.86 0.10 0.80 0.82 0.42 0.11 0.24 0.66
CCCma (Avg) 0.96 1.68 0.44 0.71 1.26 0.10 1.03 0.99 0.51 0.12 0.24 0.72
CCCma (90%) 1.06 2.09 0.54 0.85 1.76 0.11 1.23 1.16 0.59 0.13 0.25 0.79
CSIRO (10%) 0.74 1.41 0.46 0.35 0.76 0.11 0.98 0.65 0.42 0.15 0.25 0.65
CSIRO (Avg) 0.84 1.63 0.63 0.42 1.11 0.13 1.13 0.82 0.51 0.17 0.26 0.71
CSIRO (90%) 0.92 1.82 0.79 0.50 1.44 0.14 1.34 0.99 0.58 0.18 0.27 0.78
GFDL-CM3 (10%) 0.72 1.13 0.61 0.68 0.78 0.10 0.63 0.52 0.45 0.14 0.21 0.66
GFDL-CM3 (Avg) 0.85 1.44 0.87 0.87 0.97 0.11 0.76 0.70 0.68 0.16 0.21 0.72
GFDL-CM3 (90%) 0.93 1.72 1.23 0.99 1.12 0.12 0.86 0.86 0.83 0.17 0.21 0.76
GFDL-ESM2M (10%) 0.86 1.64 0.76 0.50 0.98 0.11 0.62 0.60 0.48 0.15 0.27 0.70
GFDL-ESM2M (Avg) 0.99 1.92 1.17 0.86 1.22 0.12 0.80 0.89 0.63 0.18 0.29 0.85
GFDL-ESM2M (90%) 1.09 2.16 1.53 1.06 1.47 0.13 0.99 1.27 0.82 0.19 0.30 0.94
MPI (10%) 0.60 1.16 0.56 0.28 0.87 0.11 0.55 0.61 0.38 0.15 0.21 0.62
MPI (Avg) 0.70 1.43 0.79 0.36 1.18 0.12 0.73 0.84 0.48 0.16 0.21 0.67
MPI (90%) 0.80 1.72 1.02 0.44 1.44 0.13 0.90 1.09 0.58 0.18 0.22 0.72
NCAR (10%) 0.75 1.38 0.58 0.41 0.93 0.09 0.71 0.57 0.52 0.11 0.21 0.63
NCAR (Avg) 0.86 1.58 0.68 0.54 1.23 0.10 0.91 0.74 0.62 0.12 0.22 0.69
NCAR (90%) 0.95 1.86 0.78 0.66 1.49 0.11 1.10 0.95 0.72 0.13 0.22 0.75
SMHI_KNMI (10%) 0.70 1.74 0.24 0.39 0.80 0.14 0.53 0.51 0.35 0.12 0.24 0.59
SMHI_KNMI (Avg) 0.81 1.98 0.25 0.43 1.13 0.15 0.65 0.80 0.40 0.13 0.24 0.64
SMHI_KNMI (90%) 0.88 2.15 0.27 0.46 1.52 0.15 0.74 0.90 0.47 0.14 0.25 0.68
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.05

Table 4.