December 2018

Multidecadal Variability in Climate Models and Observations

Alex Carl Oser
University of Wisconsin-Milwaukee

Recommended Citation

Oser, Alex Carl, "Multidecadal Variability in Climate Models and Observations" (2018). *Theses and Dissertations*. 2004.
https://dc.uwm.edu/etd/2004

MULTIDECADAL VARIABILITY IN CLIMATE MODELS AND OBSERVATIONS

by

Alex Oser

A Thesis Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Master of Science
In Atmospheric Science

at
The University of Wisconsin-Milwaukee
December 2018

ABSTRACT

MULTIDECADAL VARIABILITY IN CLIMATE MODELS AND OBSERVATIONS

by

Alex Oser

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Sergey Kravtsov

Climate change attribution and prediction using state-of-the-art models continue to garner an ever-growing focus amongst both the scientific community and public alike. Recent analyses showing discrepancies in the structure of modeled and observed decadal climate variability (DCV), therefore, have engendered efforts to not only diagnose the dynamics underpinning observed DCV, but also to characterize the behavior of DCV within climate models. In this thesis, we employ Multichannel Singular Spectrum Analysis (M-SSA) to show that while the DCV signal in observations is best described as a coherent oscillation with complex propagation across the globe, modeled DCV lacks this structure altogether. Specifically, the modeled DCV has a considerably smaller magnitude than its observed counterpart, and tends to exhibit simpler spatiotemporal behaviors. In particular, within the vast majority of models, the DCV structure is best characterized either by globally synchronous, quasi-oscillatory patterns lacking propagation, or, secular trends punctuated with weak, oscillatory-like signals. Both observed and simulated DCV has the largest magnitude in the polar regions. However, the observed anomaly propagation suggests Atlantic control, whereas it is the Arctic that appears to be setting the tone for globally averaged variability in most model runs. Broadly, these results confirm contrasting DCV structure within models and observations, while identifying some qualitative commonalities between the observed and simulated quasi-oscillatory behavior within a few model simulations, thus providing important clues for further DCV research.

TABLE OF CONTENTS

LIST OF FIGURES

analogous figures, except for observations, are shown below them in **(c)** and **(d)**, respectively.

Figure 19 (Page-48): Analogous to Figure 12 except for MIROC5 run 1 **(a)** and CanESM2 run 3 **(b)**. Both of these model runs fall within category 3 (“non-pair”) a linear trend punctuated by oscillations and a mixture of two-or-more oscillatory signals, respectively. The analogous figures, except for observations, are shown below them in **(c)** and **(d)**, respectively.

LIST OF TABLES

1. Introduction

We live in a time when there is never a dearth of information regarding our planet's climate available to us, regardless of whether we seek it or not, and perhaps more importantly, regardless of its accuracy and scientific acuity. A recent example, which, aided by media and social-media sources likely made its way quickly across the globe, would be the Intergovernmental Panel on Climate Change's (IPCC) Special Report finalized in October of 2018, stating, "Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a *likely* range of 0.8°C to 1.2°C. Global warming is *likely* to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate. (*high confidence*)" (IPCC 2018). It is the right of each individual who comes in contact with this information to digest it as they see fit, while it is a charge to the scientific community, and a source of motivation, to categorize it as another piece of evidence in the quest to better understand the behaviors of our climate system in both the short and long term. One such effort underway in the climate community is the pursuit of knowledge regarding our climate's variability on time-scales from a few decades to approaching one-century (decadal climate variability [DCV]). In particular, the predictability of future states of the climate relies on, amongst many other things, our understanding and proper adoption of ensemble-climate-models to study DCV events (anomalies), including but not limited to their origins, evolution across the globe, and interplay with external forcing, whether natural or anthropogenic (Cassou et al. 2018). Many strides have been made within the DCV arena, especially when considering *observed* surface temperature variability (both over land and over oceans) and our increased understanding of how our planet's oceans absorb and distribute heat on decadal time-scales in both a regional sense, as well as globally (Yan et al. 2016). These large-scale, low-frequency variability phenomena, such as the Atlantic Multi-decadal Oscillation (AMO: Buckley & Marshall 2016; Yeager & Robson 2017) and the Pacific Decadal Oscillation (PDO: Newman et al. 2016), while sometimes described in regional terms, tend to “imprint” themselves on the global-scale through a network of teleconnections, representative of either an atmospheric forcing of, or response to, such oceanic variability (Cassou et al. 2018). While we continue to advance our understanding of the physical mechanisms surrounding oceanic DCV (Danabasoglu et al. 2016; Hedemann et al. 2017; Robson et al. 2012; Ruprich-Robert & Cassou 2015; Smith et al. 2016; Swingedouw et al. 2015; Zhang et al. 2013), one cannot lose sight of the importance for our current and yet developed global climate models (GCMs) to also accurately display both regional and global DCV patterns.

Despite a recent surge in the amount of available climatic *observations*, historical records exhibit a relative scarcity temporally, and an overall scarcity spatially, with an emphasis over the oceans and especially over the Southern Hemisphere (Deser & Phillips 2017). This, as well as the overarching complexities of the climate system, lead to shortcomings in our state-of-the-art global-scale climate GCMs, which have to use ad hoc approximations for the physical processes they don’t explicitly resolve. Multi-scale interactions ultimately responsible for the Earth’s weather and climate lead to unintended, and due to current memory and processing inadequacies, unavoidable, model errors, as parameterizations of subgrid-scale processes certainly feedback on DCV (Loehle 2018; Stevens & Bony 2013; Booth et al. 2012; Evan et al. 2013; Martin et al. 2014, Yuan et al. 2016; Brown et al. 2016). Cloud processes and their effect on the surface energy budget are one such example where model parameterizations lead to surface variability errors, and thus impact the GCM's ability to accurately capture DCV (Arakawa 2004; Park & Bretherton 2009; Zhang & Mcfarlane 1995). Thus, while we must continue to monitor, research, and diagnose DCV from *observations* and their resulting global impacts in the near term (see Wang et al. 2012; Schubert et al. 2004; Chylek et al. 2014; van Dijk et al. 2013), we must not lose sight of the importance of our GCM's limited ability not only to accurately hindcast past DCV events, but also replicate, at least in a loose sense of the term, the spatiotemporal structure of DCV, which is evident in *observations*, and has recently been borne out in efforts to characterize *observed* DCV structure (Kravtsov et al. 2018).

The aim of this thesis is to accomplish exactly such a characterization of the dominant DCV modes within the Coupled Model Intercomparison Project, Phase 5 (CMIP5: Taylor et al. 2012). We already know that differences exist within the spatiotemporal structure as well as magnitude of *observed* DCV when compared to *modeled* DCV (Kravtsov & Callicutt 2017; Swanson et al. 2009; Knutson et al. 2016; Kravtsov 2017; Qasmi et al. 2017). Therefore, it is imperative we document the precise space–time structure of *modeled* DCV, with a short-term goal of comparing this structure to the one characterizing *observed* DCV, and a long-term goal of spurring the next generation of GCM's to more accurately reflect *observed* DCV in the past as well as the future.

The remaining contents of the thesis are organized as follows: section 2 provides a description of the data sets and outlines the methodologies used to analyze *modeled* DCV therein, as well as synthesizes the data and methods used to obtain *observed* DCV;

a descriptive analysis of DCV within *observations* and *models* follows in sections 3 and 4, respectively; section 5 then concludes the analysis with a discussion regarding the similarities and differences between *observed* and *modeled* DCV, culminating in suggestions for future work.

2. Data sets and methodology

2.1. CMIP5 model simulations and identification of their multidecadal internal variability

Our analysis of *modeled* DCV used particular model ensembles from the CMIP5 Project (Table 1), which were previously analyzed in Kravtsov (2017). This dataset is comprised of 111 total model runs from seventeen unique GCMs, each with three, five, six, or ten independent, “long-term” simulations, containing monthly historical surface atmospheric temperature (SAT) data interpolated on a 2.5º by 2.5º grid using cubic splines, and spanning the time period from 1880 to 2005. Kravtsov and Callicutt (2017) and Kravtsov (2017) showed that the single-model ensemble means for ensembles with three or more simulations provide a fairly accurate estimate of the forced signal in these ensembles.

We focus here on multidecadal climate variability, which may contain both forced signal and low-frequency internal variability. To isolate the secular variability in SAT, Kravtsov et al. (2018) developed an original methodology based on a combination of Multichannel Singular Spectrum Analysis (M-SSA; Moron et al. 1998; Ghil et al. 2002) and classical optimal (Wiener) filtering, in which the contribution of the high-frequency, internal noise to each M-SSA mode is estimated using multi-scale linear inverse modeling (LIM) (Penland 1989; Penland 1996; Penland & Sardeshmukh 1995; Kravtsov et al. 2005; Kravtsov et al. 2017; Jeffrey et al. 2013), and then removed. The remaining filtered, low-frequency (multidecadal) variability within each model run will hereafter be referred to as the *secular signals* of the model. While the methods which undergird these filtering procedures are outside the scope of this paper, Kravtsov et al. (2018) showed that these secular signals so isolated are statistically robust and can also be estimated with relative accuracy using more standard time-filtering methods.

After the secular signals in each *model* simulation have been obtained as described above, we further computed the estimate of the *secular internal variability* (again, for each *model* simulation) by forming differences between the secular signals from individual *models* and their corresponding *model* ensemble-mean secular signal (representing an estimate of the forced signal for this *model*). In the same way, we can obtain estimates of the internal variability in *observations* (Kravtsov et al. 2018) by linearly subtracting the secular signals from individual CMIP5 *model* simulations (which can be shown to be dominated by the forced signal) from the *observed* secular signal. Kravtsov et al. (2018) showed that the *observed* secular internal variability so computed has a particular structure suggestive of a global multidecadal oscillation. These authors also provided evidence that this structure is absent from the internal variability of CMIP5 *model* simulations. In the present work, we aim to describe in detail the phenomenology of the multidecadal internal variability in CMIP5 *models* and identify their key distinctions from the *observed* variability.

2.2. Analysis of space-time structure of internal variability in models and observations using M-SSA

To analyze the space–time structure of the observed and modeled internal secular variability, we used M-SSA analysis (Moron et al. 1998; Ghil et al. 2002), which is a multivariate extension of the Singular Spectrum Analysis (SSA: Broomhead & King 1986; Vautard & Ghil 1989; Vautard et al. 1992), and to a broader extent, that of the more standard principal component analysis (PCA) or Empirical Orthogonal Function (EOF) analysis (Monahan et al. 2009). In this way, we are able to analyze the underlying structure of the secular signals in the available (modeled and observed) climatic time series, all of which are relatively short and noisy (Broomhead & King 1986). The M-SSA analysis is, in fact, the standard EOF analysis applied to the so-called trajectory matrix constructed by appending to the original multivariate time series under consideration (in our case, time series of the gridded surface temperature data) time-lagged copies of itself. In contrast to the standard EOF analysis, which finds the spatial patterns with pronounced variance, the resulting decomposition thus carries the information about the dominant modes of variability characterized by pronounced correlations in both space and time, which can provide a compact description of the time development associated with such variability.

The Singular Value Decomposition (SVD) of the trajectory matrix yields eigenvectors which are analogous to the Empirical Orthogonal Functions (EOFs), but, due to the temporal augmentation in the trajectory matrix, are called space-time EOFs (ST-EOFs). In a similar sense, space-time PCs (ST-PCs) can then be obtained by projecting the trajectory matrix onto ST-EOFs. The variances of the ST-PCs are equal to the M-SSA eigenvalues, and the M-SSA modes are sorted in the order of decreasing variance, so that the leading modes describe most of the variance. Finally, the variability associated with each M-SSA mode can be reconstructed in the original physical space to yield the corresponding Reconstructed Components (RCs). The RCs are narrowband filtered versions of the original input time series, with filter weights determined data adaptively within the M-SSA procedure. The RCs can be correlated, but the sum of all RCs is equal to the original time series in each *channel* (e.g., for each grid-point time series of the input data).

The number of lagged copies *M* to be used within M-SSA — the so-called embedding dimension — is an adjustable parameter. In the standard M-SSA application for identification of the oscillatory signals in climatic time series, *M* is chosen based on the expected range of periodicities to be detected. Each M-SSA mode can be associated with its own dominant frequency, determined by best harmonic fit to the corresponding T-EOF (see below). The pair of M-SSA modes with similar variance and frequency whose ST-EOFs and ST-PCs are in quadrature identifies an oscillation. Note, however, that in the context of the secular signals which vary, by definition, on a time scale comparable with the length of the available time series, the periodic nature of a given signal cannot be established in principle, even if it is associated with an M-SSA pair as defined above and has oscillatory-looking RCs. Therefore, here, we chose *M* by requiring the M-SSA spectrum to be optimally peaked in the sense that the leading modes would be most statistically distinct from the noise background (these modes, when combined, would represent our dominant signal) *and* describe the largest fraction of the total data variance (thus maximizing the signal-to-noise ratio). This leads to the choice of *M*=65 in the M-SSA analysis of the data–model secular differences.

Prior to the M-SSA analysis, we compressed the original gridded input time series using PCA. We found that, in the most extreme case, twenty-two leading PCA-modes were needed to account for at least 99% of the variance within the secular signal data. As such, we chose to compress the input of our M-SSA to the leading twenty-five PCs, which therefore account for nearly 100% of the variance within all of the model runs, and act as the twenty-five channels within the M-SSA. We then used the resulting ST-EOFs and ST-PCs to obtain the RCs for each realization of the estimated secular variability (modeled or observed) and transformed them back into physical space to approximate secular signals on the original grid for plotting and analysis.

To analyze the frequencies which dominated a particular M-SSA mode, we regressed the ST-EOFs (representative of each M-SSA mode within a given multi-channel input time series) onto a set of harmonic predictors (sines and cosines), while varying both channel and frequency. The frequency range chosen was 1/500=0.002-yr^{-1} through 1/10=0.1-yr^{-1}, representing frequencies in the range of 10 to 500 years, while the process of choosing channels proved to be a bit tricky. Ultimately, our choice of which channels to consider within the regression process was based off of two measures of the signal strength within each channel for a given model. The first was the fraction of the variance accounted for by a given RC within each PC channel. These ratios were then sorted in decreasing order, as all would be less than one by definition, and the channel whose variance most closely matched the original PC variance was considered the “strongest” channel. In a similar manner, the variance of each channel within a model’s

RCs were normalized by the sum of the variances of all channels within that model's RCs, once again giving us a measure of strength for each particular channel within the set of twenty-five channels for a model's RCs. These were also sorted in decreasing order, and the ranks of both strength measures where then summed, with the channel associated with the smallest sum then used in the regression process to determine the dominant frequency of the M-SSA mode under consideration. In 102/111~92% of the cases (out of 111 model simulations considered), the first channel was used, in 8/111~7% the second channel was used, and in 1/111~1% the third channel was used for that purpose.

3. Observed DCV – The Global Stadium Wave (GSW)

This section summarizes the results of Kravtsov et al. (2018) analysis of the observed and CMIP5 simulated secular variability. We first verified our methodology for obtaining the secular SAT variability using Wiener filtering (Section 2.1) using the simulations from the Community Earth System Model (CESM) Large Ensemble Project (LENS: Kay et al. 2015) (see Figure 1 for an example of M-SSA based Wiener filtering: Each non-stationary (secular) SAT signal was defined to be associated with the part of the SAT singular spectrum (inferred via M-SSA) that cannot be simulated by stationary linear inverse trained on pre-low-pass filtered data. The filter weights were derived via computing the signal-to-noise ratio of each M-SSA mode, and the weighted M-SSA decomposition is transformed back to physical space to reconstruct the part of variability associated with the signal). Forty available simulations of the twentieth century variability reflect a common forced signal summed with independent realizations of internal climate variability. Each SAT simulation was filtered as described above. We then compared the non-stationary signals to estimated forced signals defined via the ensemble average of the surface temperature over all of the 40 simulations. The reconstructed non-stationary signal closely resembles the low-frequency forced response of the CESM model but fails to capture the short-term temperature response to episodic volcanic eruptions (Figure 2). The secular *internal* variability in LENS simulations is, therefore, relatively small, consistent with a recent study (Bellomo et al. 2018).

In the main part of our analysis, we considered 17 ensembles of the CMIP5 models with three or more historical realizations (totaling 111 simulations), as well as the gridded surface temperature product from NOAA's twentieth century reanalysis (20CR; Compo et al., 2011) in lieu of the observed SATs. Although the number of realizations in these models is much smaller than in the CESM LENS ensemble, for example, it is still sufficient to evaluate contributions from forced signal and internal climate fluctuations to these models' secular variability (Kravtsov & Callicutt 2017; Kravtsov 2017; Frankcombe et al. 2018). Similar to LENS simulations, the non-stationary signals inferred from CMIP5 models capture the low-frequency forced signal less the effect of volcanic aerosols; however, when considered in aggregate, they reflect a larger spread of possible secular signals due to incorporation of model uncertainty (see section 4). On the other hand, the dynamical structure of the secular signal in observations is richer than that in the models in the sense of being represented by a larger number of significant M-SSA modes (see section 4).

Given that our estimated secular signals in CMIP5 simulations primarily reflect the forced response of CMIP5 models, it makes sense to linearly subtract them from the observed secular signal to study the part of the observed secular variability unaccounted for in CMIP5 simulations (Steinman et al. 2015; Kravtsov & Callicutt 2017; Kravtsov 2017); this part can be viewed as, and will henceforth be referred to as, an estimate of the "observed internal variability." We can also subtract the individual model ensemble-mean secular signals from all of this model's simulations to define the internal component of the secular signal within each simulation. The M-SSA analysis of the observed internal variability identifies a pronounced pair of M-SSA modes altogether absent from the simulated internal variability (Fig. 3, left). The reconstruction of this pair of modes for regional climate indices (Fig. 3, right) reveals an oscillatory-like, multidecadal signal propagating across the climate index network; a so-called *stadium wave* (Wyatt et al. 2012), which we will refer to as the *Global Stadium Wave* (GSW). The pairs of M-SSA eigenmodes with similar magnitudes and time scales may indeed indicate the presence of a quasi-oscillatory mode (Ghil et al. 2002) in the data; in the context of the secular signals, which have time scales comparable to the length of the data record, the periodicity of such a signal cannot be verified, but the propagation of the anomalies in space in the course of the oscillation can still be established with statistical significance (Kravtsov et al. 2014).

The order of indices in the sequence of Fig. 3 (except for GMO) is chosen based on the visual analysis of the SAT anomaly propagation over a time period between 1921 and 1963, which roughly spans half of the oscillation period (Fig. 4). In year 1921, the oscillation is in its cold phase (Fig. 3, middle), with the exception of four major positive SAT anomaly spots: west of the Weddell Sea, in the eastern equatorial Pacific, as well as over the central US and Greenland. The development of an oscillation starts with an emergence of a positive SST anomaly in the North Atlantic (1921–30), which subsequently expands and grows as positive SST anomalies in the North and Southwest Pacific (1933–1942), then the Southern Ocean and Antarctica (1941–1957) and, finally, over the Arctic (1960–1963), at which point the oscillation arrives at its positive phase throughout the world (less four major negative SAT anomaly regions roughly at the same locations as their positive analogs 40 years previous).

As mentioned above, the phasing of the derived indices in the global stadium wave is consistent with the earlier work of Kravtsov (2017), which utilized a limited subset of Northern Hemisphere climate indices, including, in addition to surface temperature indices, the sea-level pressure based atmospheric indices, in particular, the North Atlantic Oscillation (NAO: Hurrell 1995; Hurrell & Deser 2009) index. This study showed that the observed NAO variability has a pronounced GSW-related component; by contrast, the multidecadal NAO variability in CMIP5 models is an order of magnitude smaller than that in observations (Fig. 5) and lacks the observed coherence with oceanic variability (figure not shown). We hypothesize that the lack of the atmospheric sensitivity to multidecadal climate signals originating in the ocean (of which the results regarding the NAO described above are just one example) is one of the main reasons behind the absence of the GSW in CMIP5 models, as these atmospheric teleconnections help the GSW signal propagate across the globe.

Despite Wiener filtering methodology applied in M-SSA based space–time phase space providing an inherently more accurate identification of secular signals and explicitly dealing with the issues associated with observational uncertainties in sparsely sampled regions of the globe (by employing space–time covariance based signal detection), the key differences between observed and model simulated climates on multidecadal time scales are so pronounced that can be easily detected using traditional time-filtering methods (Kravtsov et al. 2018).

These authors also considered the sensitivity of the GSW to the reanalysis uncertainties, by repeating the entire analysis procedure using two versions of the 20CR reanalysis, as well as a more recent ERA-20C reanalysis (Poli et al. 2016). The GSW reconstructions using different reanalyses show consistent behavior over most of the low- and mid-latitude World Ocean. The GSW variability over land is consistent between 20CRv2 and 20CRv2c reanalyses, but is entirely different in the ERA-20C reanalysis. Finally, the details of the GSW in the coastal Southern Ocean and Arctic region are also reanalysis dependent. Overall, the GSW space–time development exhibits consistency between the two versions of the 20CR reanalysis, except for the behavior over the narrow strip in the coastal Southern Ocean and a decadal shift in the Arctic component of the GSW. The 20CR and ERA-20C reanalyses are generally consistent over low-to-middle latitude oceans, but exhibit much larger differences over land and in polar regions. All three reanalyses thus identify the GSW emanating from the North Atlantic region and spreading over the globe via a combination of oceanic and atmospheric teleconnections, but the response over land is entirely different in ERA-20C. However, either version of the GSW is not reproduced in any of the CMIP5 simulations considered.

4. Modeled-vs.-observed DCV

Using the 111 unique CMIP5 estimates of model secular variability, as well as the corresponding estimates of the observed internal variability, we begin to explore DCV in models, and the differences compared to observed DCV, at its roots: within the *magnitude* and within the *structure* of this variability.

4.1. M-SSA based characteristics

Both of these properties are already reflected in the M-SSA spectra of the observed and model simulated secular signals. The leading M-SSA pair associated with the observed secular variability is shown in Fig. 6 in terms of its frequency–variance characteristics. It can be considered a pair as both the vertical (variance) and horizontal (frequency) error bars overlap (these error bars were computed in terms of the standard spread of the corresponding quantities over 111 estimates of the observed internal secular variability, each associated with the subtraction of the forced signal estimate from available model simulations). Figures 7–9 are analogous to Fig. 6, but show M-SSA frequency–variance spectra for the leading 10 M-SSA modes (as opposed to the leading two modes alone), for observations (Fig. 7a), and a slightly extended set of leading modes for all of the model simulations considered (all other panels — see figure captions and legends). The main feature shared by all of the M-SSA spectra is that the leading two M-SSA modes stand out of the rest of the spectrum in terms of their variance, so that these two modes describe the dominant structure in both observations and all model simulations, whereas the rest of the M-SSA modes represent the background noise with little-to-no discernable structure. This greatly simplifies our analysis by allowing us to concentrate on these two modes only to characterize the differences between the observed and simulated secular variability. The second important property is that the leading two M-SSA modes in the vast majority cases are the lowest-frequency modes with periods exceeding 30 yr. This is perhaps not surprising, since the Wiener pre-filtering of secular variability indeed concentrated on the multidecadal time scales.

To compare and contrast the magnitude of modeled versus observed variability, we thus begin by looking no further than the spread of the modeled variances of the first two M-SSA modes themselves (for observations and model simulations). Broken down into quartiles, we have four variance ranges: (i) low variance: [0,821]; (ii) medium variance: [822,1313]; (iii) high variance: [1314,2407]; (iv) very high variance: [2408,+inf). When compared to the same values for observations, the analogous quartile ranges are given by: (i) [0,8444]; (ii) [8445,10577]; (iii) [10578,13825]; (iv) [13826,+inf). Thus, collectively, the first three quartiles of the modeled variances easily fall within the first quartile of the observed variance when considering the leading pair only. When comparing the spread of the magnitudes of the variance within observations (Figure 7(a)) to those within the models (Figures 7(b-f) and 8-9(a-f)), only two models appear to have variability within the leading pair that rivals that in observations, and those are CSIRO-MK3-6-0 (Figure 7(e)), and GFDL-CM3 (Figure 8(a)). This once again quantitatively positions modeled variability well outside the norm when compared to observed variability and is further evidence that the complexities of said variability are rarely captured by our current state-of-the-art GCMs.

As previously mentioned, each of the 111 modeled leading modes of variability was compared to each of 111 observed leading modes of variability (111x111=12321 total comparisons) and only two models (CSIRO and GFDL-CM3) contained runs with a greater magnitude of variability compared to that of individual estimates of observed internal variability. More specifically, only three total model runs (CSIRO runs 4 and 5;

GFDL-CM3 run 5) showed magnitudes of variance higher than that of observations, with a grand total of slightly above 0.3% of all model runs showing higher variability in both of the leading modes (CSIRO run 4 > observations in 6 cases; CSIRO run 5 > observations in 27 cases; GFDL-CM3 run 5 > observation in 6 cases; 39/12321~0.3%). Thus, while these model runs are extreme outliers, we should look to them, as well as other models that behave similar to them, to guide our quest to better characterize modeled DCV and compare it against the observed DCV. To do so, we will visualize the structure of the variability associated with the leading two M-SSA modes, first in the phase space of the SAT EOFs (within the 25 channels used as inputs to the M-SSA analysis: section 4.2) and then by plotting the associated RCs in the physical space (section 4.3).

4.2. The structure of the observed and modeled variability in EOF phase space

Figure 10 shows the ensemble average ratios of variance between the Reconstructed Components (RCs) of the leading pair of M-SSA modes, and the variance of the leading twenty-five pre-filtered PCs used in M-SSA, for both models and observations. This can be thought of as a proxy for the *structure* of the variability, as it quantifies how well each channel within the leading M-SSA modes captured the variability within the PCs. In a general sense, only the leading channels should be considered, as this ratio degrades quickly, with both models and observations explaining less than 20% of the variance at or beyond the fifth channel. Focusing on the first two channels, we note a few distinctions between models and observations. First, the loss in variability explained between channels one and two reduces much more slowly in observations than it does in models. In fact, the variability explained by channel-two in the observations is larger than the variability explained by channel-one in the models, and the same can be said when comparing channels three and two of the observations and models, respectively. This alludes to the space-time structure in observations being more complex than it is in models, which we will later show to be the case. Specifically, the complex oscillatory patterns and propagations of observations necessitate the use of more channels in order to describe the variability within. Models, on the other hand, tend to lack these oscillatory and propagational complexities, tending more towards synchronous, low-frequency secular trends, mimicking spatial red-noise patterns with space and time memory. Second, the spread in variability amongst the first two channels is much less in observations compared to models. This further illustrates the complex yet distinct structures which must dominate in observations, as both channels one and two not only explain a relatively large amount of variability compared to models, but also do so in a more consistent fashion, leading to a relatively small amount of spread when compared to models.

4.3. The structure of the observed and modeled variability in physical space

4.3.1. Categorization of modeled DCV by frequency

In this section, we aim to elucidate the salient differences in the structure of the observed and model simulated variability. In order to accomplish this task, we divided the models into three separate categories based on the range of frequencies within the leading two M-SSA modes, while we largely ignored the differences in their variance. These groups were:

(1) the ultra-low-frequency (ULF) signals, for which at least one of the two leading M-SSA modes has a frequency less than 0.008 yr^{-1};

(2) the 'pairs', for which both leading M-SSA modes have similar frequencies larger than 0.008 yr^{-1}; and

(3) 'non-pairs', with different dominant frequencies both larger than 0.008 yr^{-1}.

The observed secular variability belongs to the group (2), which is most likely to produce oscillatory-looking signals. On the other hand, the three model runs that have most intense secular variability approaching in magnitude to the observed variability fall within categories (1) [CSIRO, Runs 4 and 5] or (3) [GFDL CM3, Run 5], and thus, first of all, have spatiotemporal structures very different from that of the observed variability (as is further illustrated below) and, second, are more likely to be associated with the low-frequency noise.

4.3.2. Ultra-low-frequency (ULF) signals

These, once again, include the model runs with the leading two M-SSA modes in which at least one of the frequencies (for mode 1, mode 2, or both) is <<0.01. These modes are typically a mixture of a secular trend and oscillatory-looking signal. The trend and oscillatory-looking signal (not an oscillation, which requires a pair) may have different spatial expressions. This group contains 65 out of the 111 CMIP5 simulations (~58.6%), and can be further divided into four approximate subgroups:

(i) In-sync oscillation-like runs (27 runs): This group contains runs which are dominated by a higher-frequency oscillatory-looking signal, and the secular trends within are relatively weak (but noticeable) in most of the regions. The amplitude of the oscillation may also be modulated in time (e.g., CCSM4, Run 4). Most of the runs from this group have in-sync variability in surface temperature all over the globe, with a few possible regional exceptions (e.g., CanESM2, Run 5; CNRM, Run 7; IPSL, Run 4). A typical feature of these (and many other) runs is that the variability in Arctic air temperature dominates (sometimes vastly; e.g. GISS-E2-Rp3, which has a strong trend confined to Arctic) compared to variability in other regions. Some of the runs, however, have comparable level of variability over Antarctica and, in few cases, other regions (e.g., CNRM, Run 4 and 8; CSIRO, Run 5; GISS-E2-Hp3, Run 3).

(ii) Runs with in-sync non-uniform trends (15 runs): These runs exhibit strong non-uniform trends, varying nearly in-sync throughout the globe, possibly with different regional amplitudes. Arctic or polar dominance are, once again, common, but there are runs with uniform amplitudes for different regional indices. A typical example of the run from this group is HadCM3 Run 4, whose regional indices from models and observations are plotted in Figure 13.

(iii) Polar dominated (often detached) trends (19 runs): These runs exhibit trends dominated by those in Arctic or Antarctic temperatures (or both at the same time). The Arctic and Antarctica can be in-phase (GISS-E2-Hp3, Run 1), out-of-phase (CSIRO, Run 4; GISS-E2-Hp3, Run 2) or phase-shifted/unrelated (GFDL-CM3, Run 3; GISS-E2-Rp3, Run 6). The rest of the world can vary in sync with polar indices (CSIRO Run 4) or completely independently (GFDL-CM3, Run 3; HadCM3, Run 10; HadGEM2-ES, Run 1).

<u>(iv) Other (4 runs):</u> These runs do not show any robust regional relationships with a discernable structure and are likely to represent stochastic ULF variability.

To illustrate the typical behavior associated with the ULF variability, we show, in Fig. 12, the reconstructions of CSIRO runs 4 (sub-category iii) and 5 (sub-category i) averaged regionally in the Arctic (AR) and Antarctic (AA) regions, Northern Hemisphere (NMO, <60ºN), Southern Hemisphere (SMO, <60ºS), and globally (GMO). We also show, for comparison, the analogous reconstructions for the observed GSW. Recall from Section 4.1, we subtract, from observations, the secular signal (presumably externally forced) from a CMIP5 model run of interest obtained by our filtering methodology (Section 2.1). The difference yielded is a distinct estimation of the internal variability within observations, of which we produce 111, from the 111 distinct CMIP5 model runs. For this analysis, we'll focus on the regions listed above as they do well to show global variability, as opposed to more equatorial and mid-latitude oceanic-regions, which can downplay the importance that polar variability has on the global scale. While many of the runs (across all categories) show strong arctic (AR) dominance, these runs show a rather comparable level of variability between the Arctic and Antarctic (a slight favorability towards the Arctic is seen in CSIRO, run 5). In order to obtain a more global perspective, as well as an understanding of the structural evolution of variability present within these two runs, figures 14 and 15 show "movies" of the global reconstructions nearly spanning the entire time period (1888-2000 as opposed to 1880-2005) and showing every eighth time step. CSIRO run 4 illustrates a well-defined "see-saw" pattern between the Arctic and Antarctic, matching the out-of-phase trends seen in the regional index plot of Figure 12 (top).

Furthermore, the amplitude and the geographical extent of these polar anomalies is more pronounced in the Atlantic sector as opposed to the Pacific sector. CSIRO run 5 (Figure 14) on the other hand, is relatively in-phase, with the Arctic slightly lagging the Antarctic, and the relative greater extent of the Arctic anomalies seen reinforces the larger magnitude of variability evident in the regional index plot of Figure 12 (bottom). As a final note, this run also contains a propagation pattern in the Antarctic, as the Atlantic sector anomalies in this region tend to propagate westerly into the Pacific sector of the Antarctic overtime, hinting at a slightly more complex space-time evolution, such as what is found in observations (Fig. 4).

4.3.3. Signals associated with M-SSA pairs

The vast majority of runs within this group (all but one; IPSL, Run 6) have a typical period from between 50-100 years. Again, we use the term “pair” for this group to describe a nearly equivalent relationship between the frequency of the two leading modes. In some cases, variances error bars overlap, lending credence to the title of “pair,” while in other cases they do not (e.g. CSIRO, Run 9). We hypothesize in the cases where variance error bars do *not* overlap, secular variability is present, as it is evident in some reconstructions (CNRM, Run 10; CSIRO, Run 7). 34 of the 111 groups fall into this group (30.6%), which can be broken into three subgroups:

(i) In-phase runs (21 Runs): Although some runs within this group still contain secular-like variability, including trends (see runs mentioned above), they also exhibit oscillatory-like patterns in which the NMO and SMO (as well as the AR and AA) are more-or-less in phase (CSIRO, Run 9; GFDL-CM2.1, Run 6; HadCM3, Run 3). We also see AR dominance in the vast majority of runs once again, with still some runs falling outside this characterization (GFDL2.1, Run7; GFDL-CM3, Run 4; HadCM3, Run 3).

(ii) Phase shifted runs (10 Runs): As the title suggests, these runs generally show phase shifts across the hemispheres, varying from a couple of decades, to some runs being completely out-of-phase (CanESM2, Run 1; HadCM3, Run 7). SMO and AA variability appears to be stronger in these runs when compared to subgroup (i) runs, however, AA variability vastly dominates some runs as well (e.g. CNRM-CM5, Run 6; MIROC5, Run 4). With the two hemispheres out of phase in general, one would assume a proper bellwether for the phase of the GMO would be to match that of the hemisphere with stronger variability. Nevertheless, it is AR variability that correlates best with GMO variability in most runs, while other runs show the GMO more in-sync with the AA (GFDL-CM2.1, Run 10), or falling somewhere in-between (HadCM3, Run 7). Lastly, a few runs show the interesting behavior of tending towards in-sync oscillations at the beginning of the time period, while becoming more pronouncedly out-of-sync during the course of the second half of the 20th century (e.g. CanESM2, Run 4). While we cannot rule out end effects as being responsible for such a shift in phases over time, the idea of GCMs producing such evolution within phase patterns is intriguing.

(iii) Other (3 Runs): Once again we include this category for completeness, with mostly secular signals mixed with quasi-oscillatory modes, but lacking any coherent oscillatory structure such as that found in observations.

We now turn to HadGEM2-ES, Run 2, as an exemplar for the M-SSA “pairs” group, highlighted by Figure 16 showing both the regional index plot for models and for observations, and Figure 17 showing this model’s temperature reconstruction “movie.” As can be seen in both the regional indices for models and the “movie,” broadly speaking, this run is in phase across the globe except for a few isolated, out-of-phase regional-foci. The period of oscillation is very close to 60-years, matching the frequency of the leading M-SSA pair in Figure 9c, and, like most models, shows a strong Arctic dominance. Looking at the regional index plot, one can see that the variability in the SMO is the weakest overall, with the smallest regional amplitude. Reconciling this using Figure 17, we see that the South Pacific and South Atlantic, which, together, make up a majority of the SMO, are examples of the out-of-phase regional foci mentioned above. This shows that while broader patterns can dominate the leading modes within models, focused, regional variability can be captured as well.

4.3.4. Signals associated with M-SSA non-pairs

This leaves us with the final, and smallest group (11/111~9.9%; GFDL-CM2.1, Run 9 contained an error, and was left out of the analysis), which are those signals associated with non-pairs. Contrary to the previous group, there is a clear mismatch in the frequencies of the leading two M-SSA modes, some of which differ by nearly 50-years (GFDL-CM3, Run 5; IPSL-CM5, Run 3), while others only differ by around 20-years (GFDL-CM3, Run 2). It should be reiterated here that while runs such as GFDL-CM3, Run 2 (Figure 7f; orange), exhibit a large mismatch in both frequency and variance of the leading pair, we still choose to observe the behavior of the leading pair as it is this pair that is dominant in observations. However, an opportunity to further investigate DCV in models exists in analyzing those modes beyond the leading pair, especially potential M-SSA pairs such as modes three and four of GFDL-CM3, Run 2.

Six of the members in this group have a similar, and rather distinct behavior, which is one of amplitude-modulated oscillation (Figure 18). In the majority of these cases, the amplitude is in-phase and decaying in time (CNRM-CM5, Run 1; CSIRO, Run6; GFDL2.1, Runs 1 and 2), with one run decaying in time while being out-of-phase (GFDL-CM3, Run 5), and the last run displaying a mix of amplitude modulation dependent on region (IPSL-CM5A, Run 3). The remaining five then fit into a broader category of “other,” differentiated by a mixture of two-or-more oscillatory signals, trends punctuated with oscillations, or for lack of a better term, noise, all of which lead to incoherent signals in regional index plots (CanESM2, Run 3; GFDL-CM3, Run 1; Giss-E2-Hp2, Run 6; MIROC5, Runs 1 and 2). It can also be seen in Figures 7-9, that the leading few modes of these particular five runs have relatively low magnitude variances within their model group, little spread in their respective variances, and differing frequencies across modes, all of which likely lead to the bizarre regional index structures seen in Figure 19.

5. Conclusion

Recently, it had been shown that the secular signals detected in observations had a coherent and complex structure. Described as the “Stadium Wave,” this structure propagated across the globe in time and lead to patterns of variability which differed from that of models in both magnitude and structure. The aim of this paper was to further investigate these differences by painting a more precise and coherent picture of DCV characteristics in models, as had similarly been done for observations.

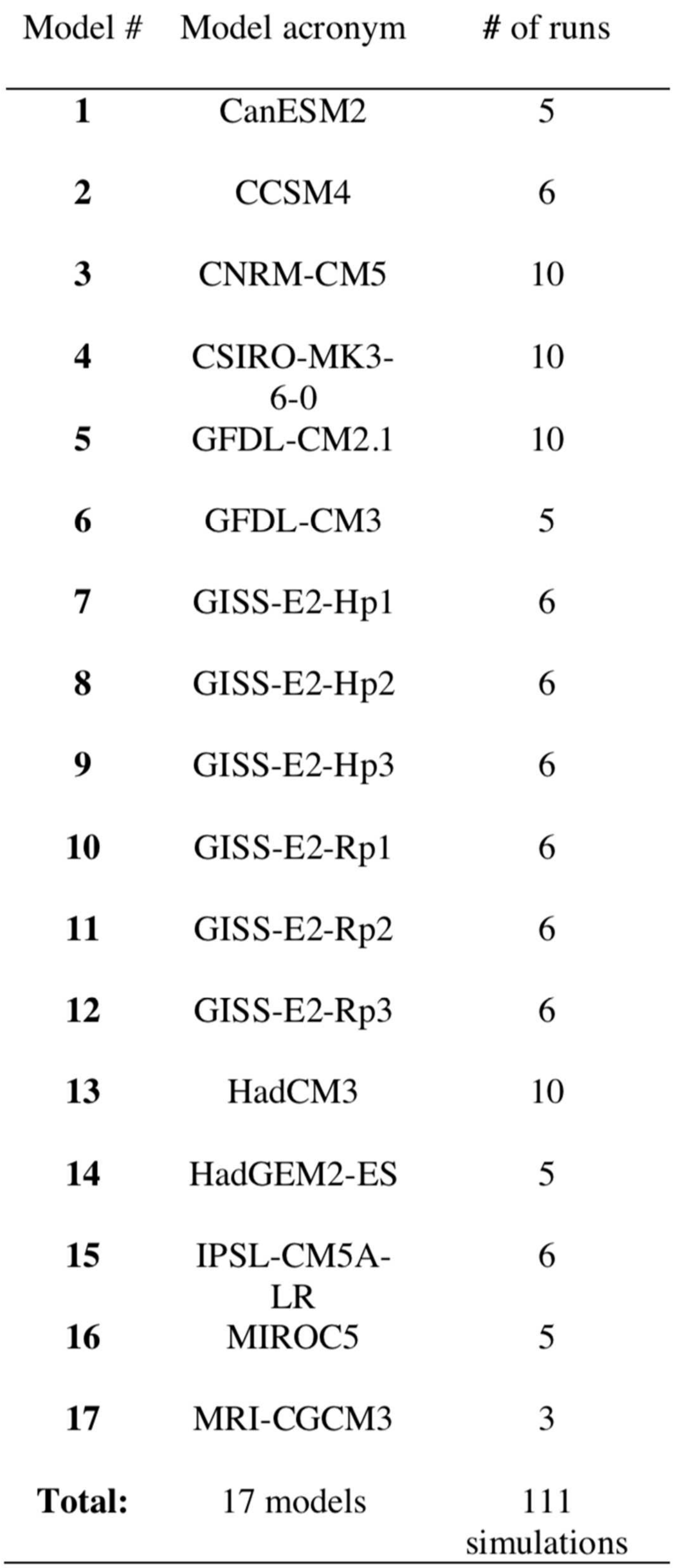
111 long-term simulations from the CMIP5 Project were used as the base set of modeled data, which were uniquely high-pass filtered in the same manner as had been done previously in the literature, in order to remove the high frequency variability inherently present in our GCMs, allowing us to focus on the secular DCV structure of interest. Furthermore, ensemble means representing the best-guess-forced-signal within a model type were subtracted from each individual model run of that type, thus yielding the secular, internal variability within an individual model run. In this same way, we also subtracted individual, secular model runs from a set of observations. By doing so, we obtained a set of 111 unique estimates of the same internal, secular variability, except this time for observation, allowing for comparison amongst the modeled and observed DCV structure.

The analysis was comprised of a combination of Principal Component Analysis for data compression, as well as the isolation of the leading modes of variability which were used as the channels in the next step of our procedure, Multichannel-Singular Spectrum Analysis, wherein we applied an embedding dimension of 65-years. We then used the resulting space-time EOF structures obtained from the Reconstructed Components of the leading two modes of variability (channels one and two) to analyze the structure of the variability in both EOF-space and physical-space, with a primary focus on characterizing these space-time structures within the models and a secondary focus of comparing our results to the observed structures described in the literature (and recreated within this analysis).

First, the amplitude and structure of variability within the models is much different than that of observations, with multiple modes of variability needed to capture the complex oscillatory and propagational behaviors found within observations, and the amplitudes of variability lagging far behind their observed counterparts. Models, on the contrary, have much of their variability isolated to the first mode alone. The RC-to-PC ratio for the second mode in models for example is close to 0.45, while the analogous mode in observations is above 0.75, and the even the third mode within observations has a greater RC-to-PC ratios (slightly under 0.5). A common theme can be seen in the PCs themselves, as the ensemble average of PCs with 1 standard deviation of error shows no overlap in the leading pair of PC modes (the leading M-SSA pair), which is a measure of statistical significance concerning the uniqueness of the variability each mode explains. Once again, the models show the opposite, with an overlap in the error bars across multiple leading modes, suggesting a lack of significant variability, and increasing the likelihood that models are behaving more stochastically than observations. This issue alone presents an opportunity for future work.

In terms of physical space, while there remain many differences which are discussed below, it should be noted from the beginning that there are in fact similarities between the temperature reconstructions. While this is a bit puzzling considering the stark differences in the variability metrics outlined above, the quasi-oscillatory nature found in many of the models is motivating. To begin, we divided the models into three groups based mostly on the frequencies found in their leading M-SSA pair, as it is this pair that dominates in explaining observed variability. These three groups were (i) Ultra-Low Frequency; (ii) M-SSA “Pairs”; (iii) Non-Pairs. The Ultra-Low Frequency group was defined by one or both of the leading modes associated with a frequency <<0.01 (100-yr periods within oscillatory-like behavior) and was the largest of the three groups with nearly 60% of models representing it. Therein, quasi-oscillatory patterns mix with secular trends to define the behavior of the bulk of the runs. To reiterate, by definition, these are not oscillations according to M-SSA, which requires the leading pair to have (nearly) the same frequency, and preferably, the variance bars to overlap. The second group, titled M-SSA "Pairs," required frequencies of the leading pair of modes to "match" (preferably within a decade of one another), however, an overlap of variance error bars was not necessary. Herein, these "pairs", while loosely defined as such by our M-SSA, mostly had periods of oscillation between 50-100 years. Given our embedding dimension of 65-years, this supports the result that, even within "pairs," the internal variability in models is secular to quasi-secular for all runs. Non-pairs make up the final and smallest group with less than 10% of the runs represented, and commonly had amplitude modulations tending towards decay in time.

Ultimately, common themes were observed across all of the groups, many of which encourage opportunities for future work. Mismatches in both frequency and amplitude of variance within the leading M-SSA pair are one such commonality already touched on but worth mentioning again. Arctic variability, and a tendency towards in-sync oscillations when oscillatory behavior was present, are two other examples, the latter of which is inconsistent with observations. Global synchronicity, in fact, suggests a need for a more focused regional analysis, and one in which future work is underway. For example, regional index plots herein focus rather broadly across hemispheres, globally, and in the poles. Moving forward, further analysis is planned for the oceanic regions, in particular the North Atlantic and North Pacific. In this same vein, the regional variability within the models was all but absent over land, save the Antarctic. All of the above further support the need for future research aiming towards improving the feedback between the oceans, atmosphere, and land within GCMs, and more accurately producing the past and future global variability they aim to explain.



Model #	Model acronym	# of runs
1	CanESM2	5
2	CCSM4	6
3	CNRM-CM5	10
4	CSIRO-MK3-6-0	10
5	GFDL-CM2.1	10
6	GFDL-CM3	5
7	GISS-E2-Hp1	6
8	GISS-E2-Hp2	6
9	GISS-E2-Hp3	6
10	GISS-E2-Rp1	6
11	GISS-E2-Rp2	6
12	GISS-E2-Rp3	6
13	HadCM3	10
14	HadGEM2-ES	5
15	IPSL-CM5A-LR	6
16	MIROC5	5
17	MRI-CGCM3	3
Total:	17 models	111 simulations

Table 1: The seventeen CMIP5 twentieth century simulations and their corresponding number of runs (111 model-runs analyzed in total)

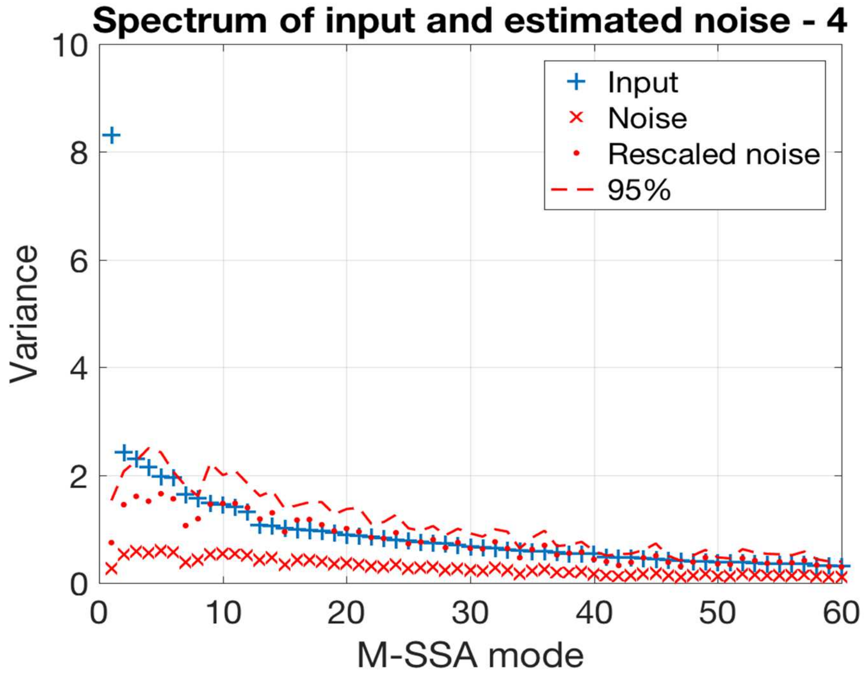


Figure 1: M-SSA spectra associated with run 4 from the LENS project. The raw spectrum of input signal (blue plus signs) matches the rescaled noise spectrum (red dots) very well in the tail of the spectrum. Only two leading input modes, however, exceed the 95th percentile of the variance associated with the stationary noise model and will be used to reconstruct the secular signal in this simulation.

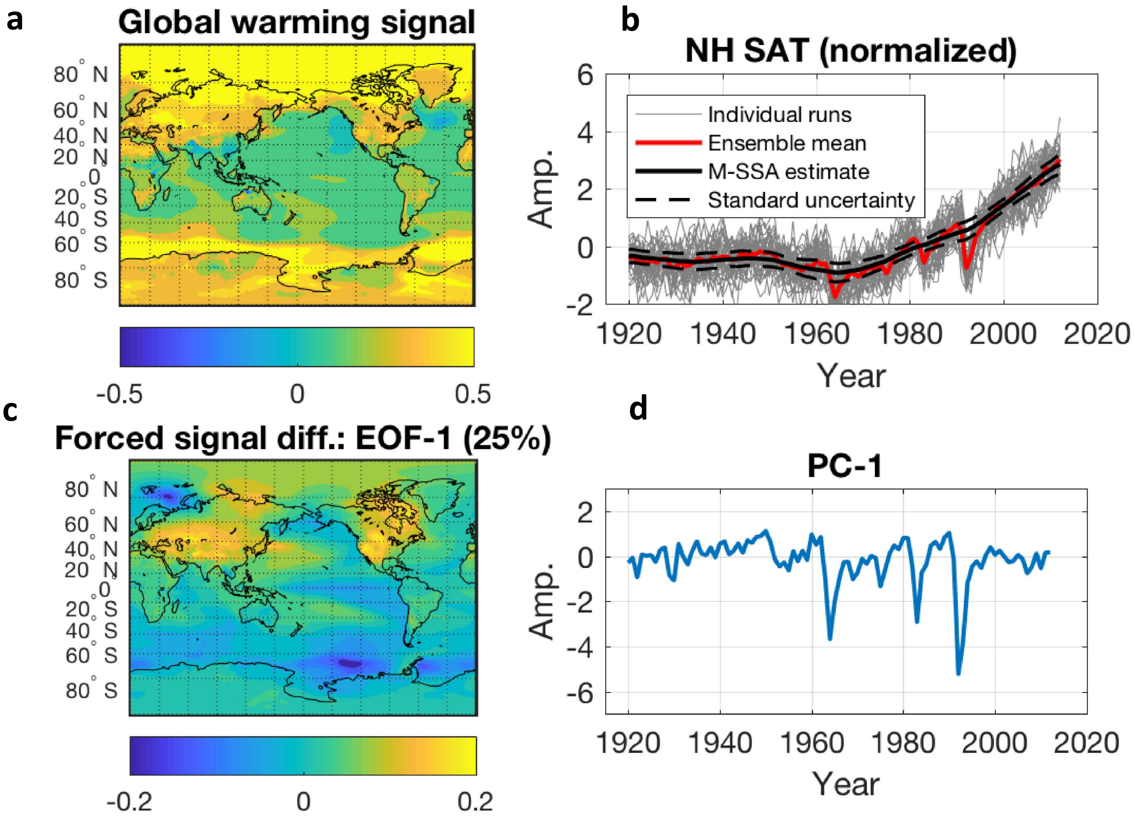


Figure 2: (a) Ensemble-mean global-warming pattern (ºC) obtained by regressing the secular SAT signals onto centered and normalized time series of their Northern-Hemisphere mean (black line in **(b)**). **(b)** Normalized time series of the Northern Hemisphere mean SAT. Individual simulations, thin grey curves; ensemble mean, red curve; ensemble mean of secular signals, solid black curve; standard uncertainty of the secular signals (over 40 estimates), dashed black curves. **(c)** and **(d)** Spatial pattern (ºC) and normalized time series of the leading mode of the difference between the estimated secular SAT signal and the CESM's ensemble-mean SAT ('true' forced signal).

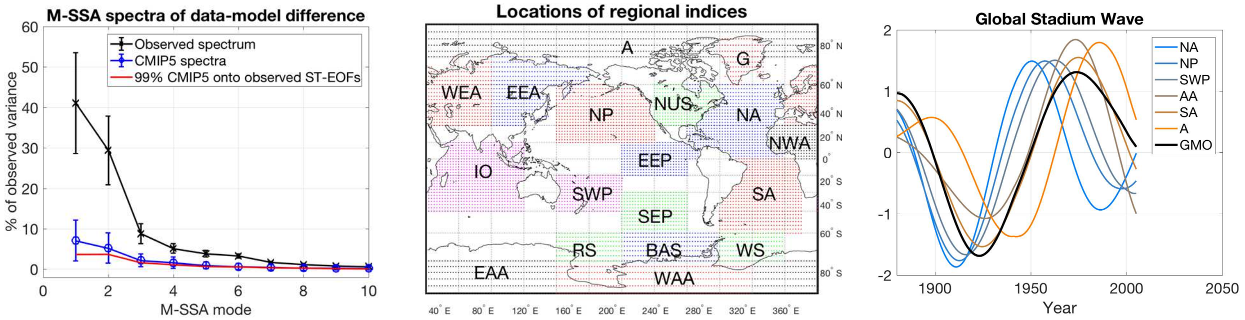


Figure 3: (Left) M-SSA spectra of observed–modeled data secular difference (black); the error bars show standard uncertainty computed over 111 estimates. Also shown are M-SSA spectra of model signals' deviations from individual model ensemble means (blue), and the 99th percentile of variances obtained by projecting the simulated signals onto the observed ST-EOFs of M-SSA analysis (red). **(Middle)** locations of regional SAT indices. **(Right)** Reconstructed time series associated with the leading M-SSA pair in select regional indices. GMO (Global Multidecadal Oscillation) time series represents the reconstruction of the global-mean temperature. All the time series are dimensionless; the actual standard deviations of A and AA indices is around 0.6K; that of all others – 0.1K.

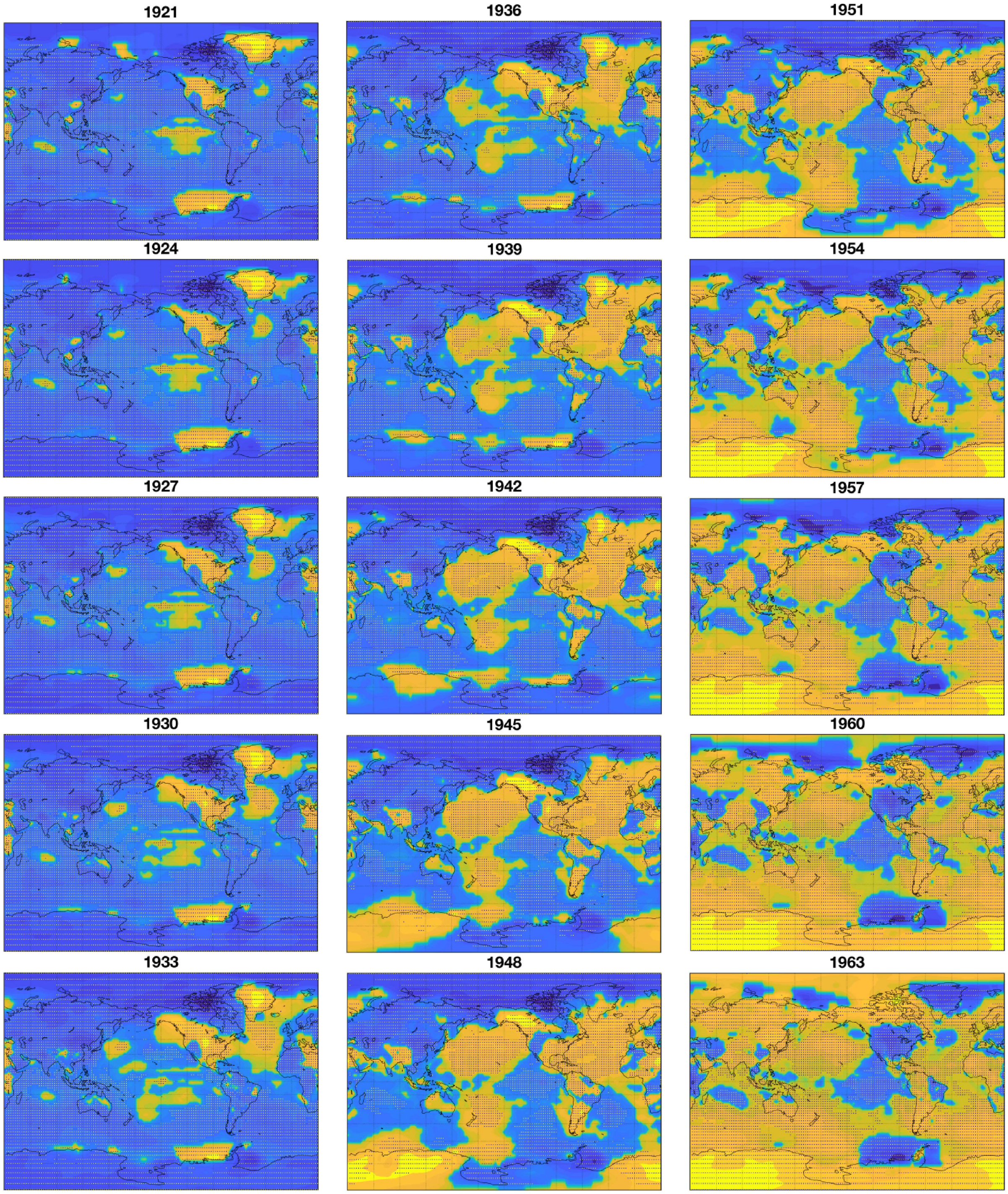


Figure 4: A 1921–1963 segment of the global stadium wave; shown are reconstructed SAT anomalies raised to the power of 1/7, which alleviates differences between SAT anomalies over ocean and over land to concentrate on the anomaly patterns and their propagation. Color axis is from –1.5 (saturated blue) to 1.5 (saturated yellow).

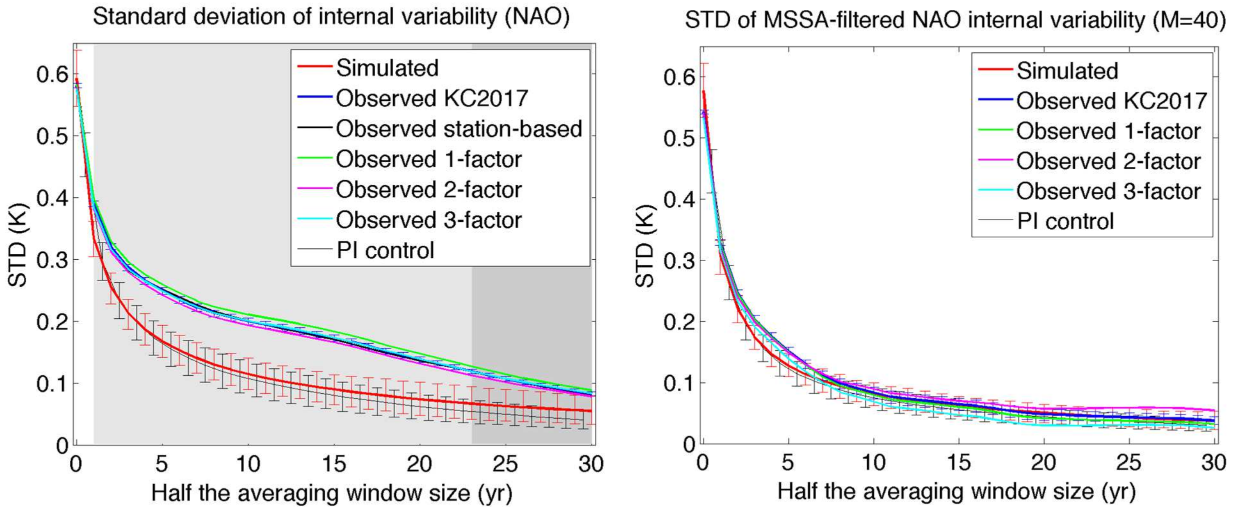


Figure 5: The standard deviation of the boxcar running-mean filter averaged NAO index, as a function of the averaging window in CMIP5 simulations (red – historical, black – control, other colors — 20CR and station-based observations). **(Left)** the results for original indices; **(Right)** the same as in the left, but with the GSW component filtered out of the observed NAO. Error bars indicate standard spread across multi-model ensemble considered. The difference between observations and model simulations are entirely dominated by the presence of the GSW component in the observed NAO data.

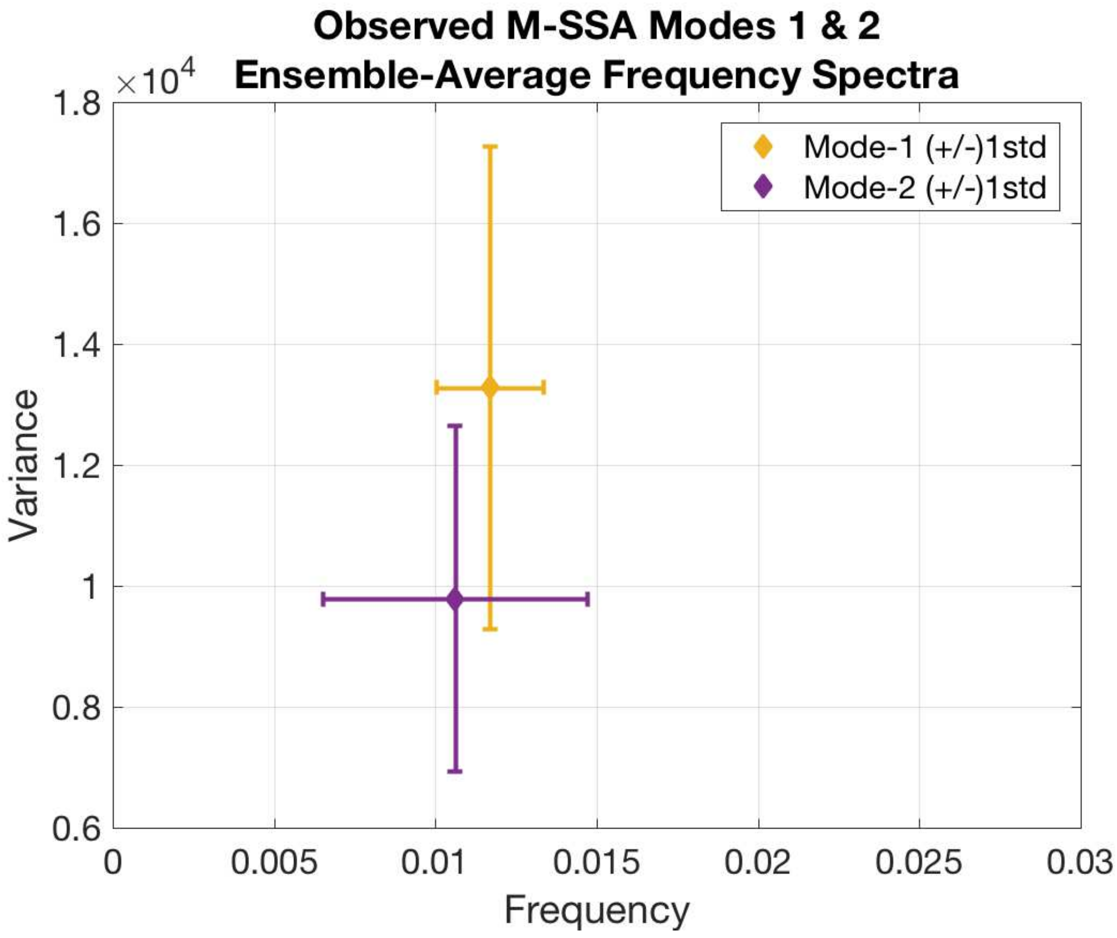


Figure 6: Variance and frequency across the 111 estimates of observed internal variability for the leading observed M-SSA pair (frequency has units: yr^{-1}).

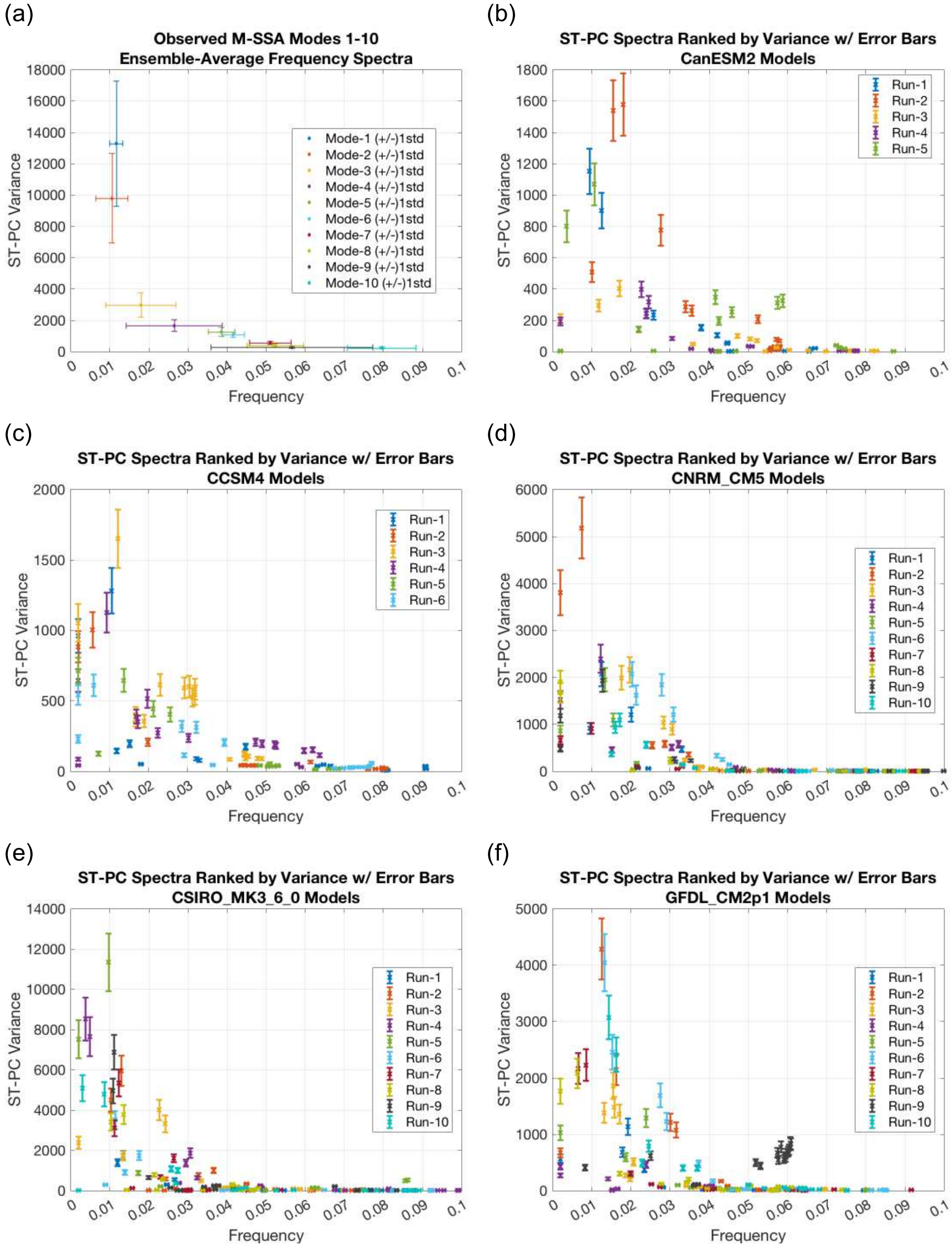


Figure 7: (a) is a recreation of figure 6 except this time showing modes one through ten. Note that modes one and two are a different color compared to figure 7; **(b-f)** Frequency spectra plots of the leading eighteen modes for each of the first five models listed in Table 1 (models 1-5), in that order. Each individual run within a model is identified by a different color, with frequency along the x-axis and ST-PC variance along the y-axis.

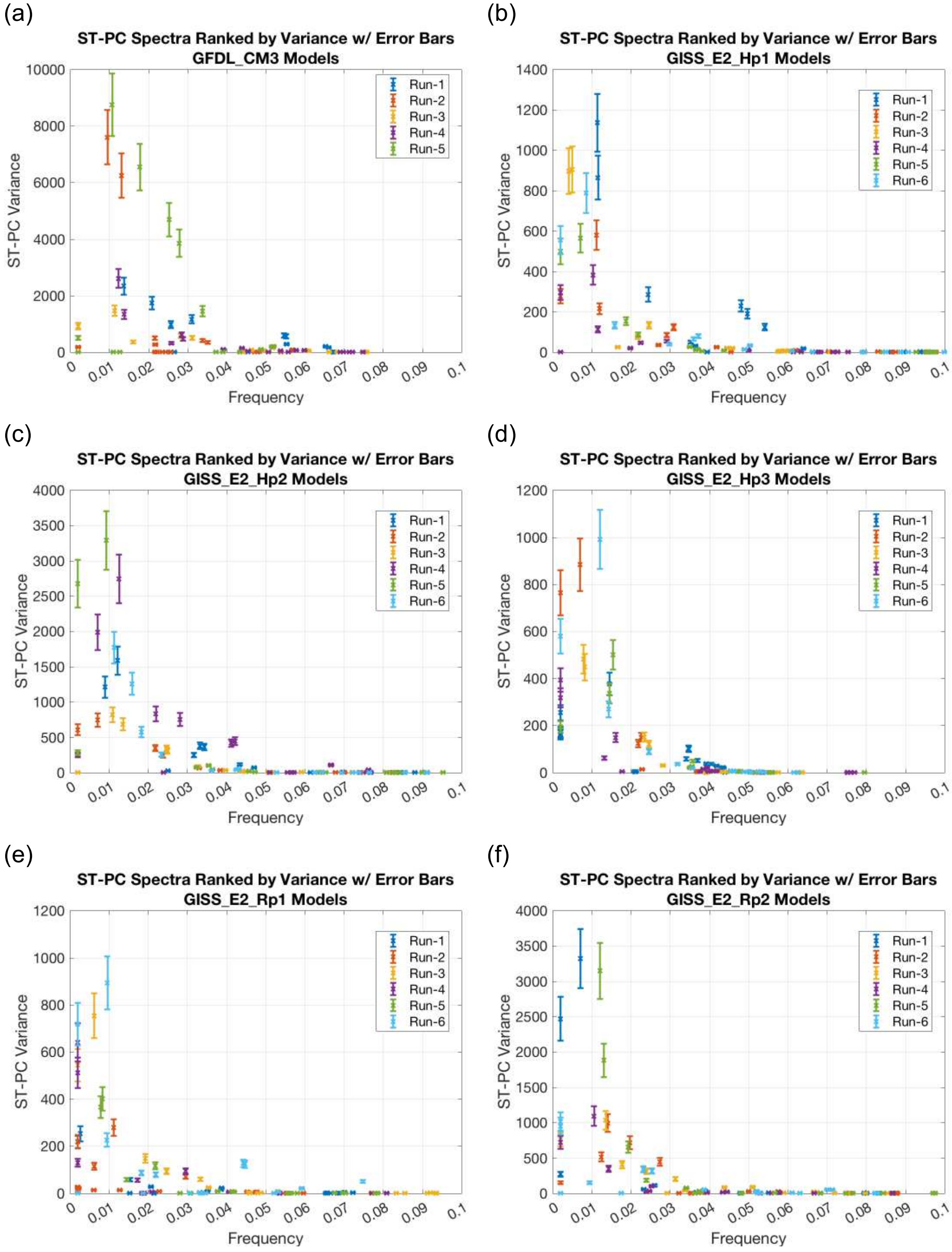


Figure 8: (a-f) Analogous to figure 7 except this time for the next six models listed in table 1 (models 6-11).

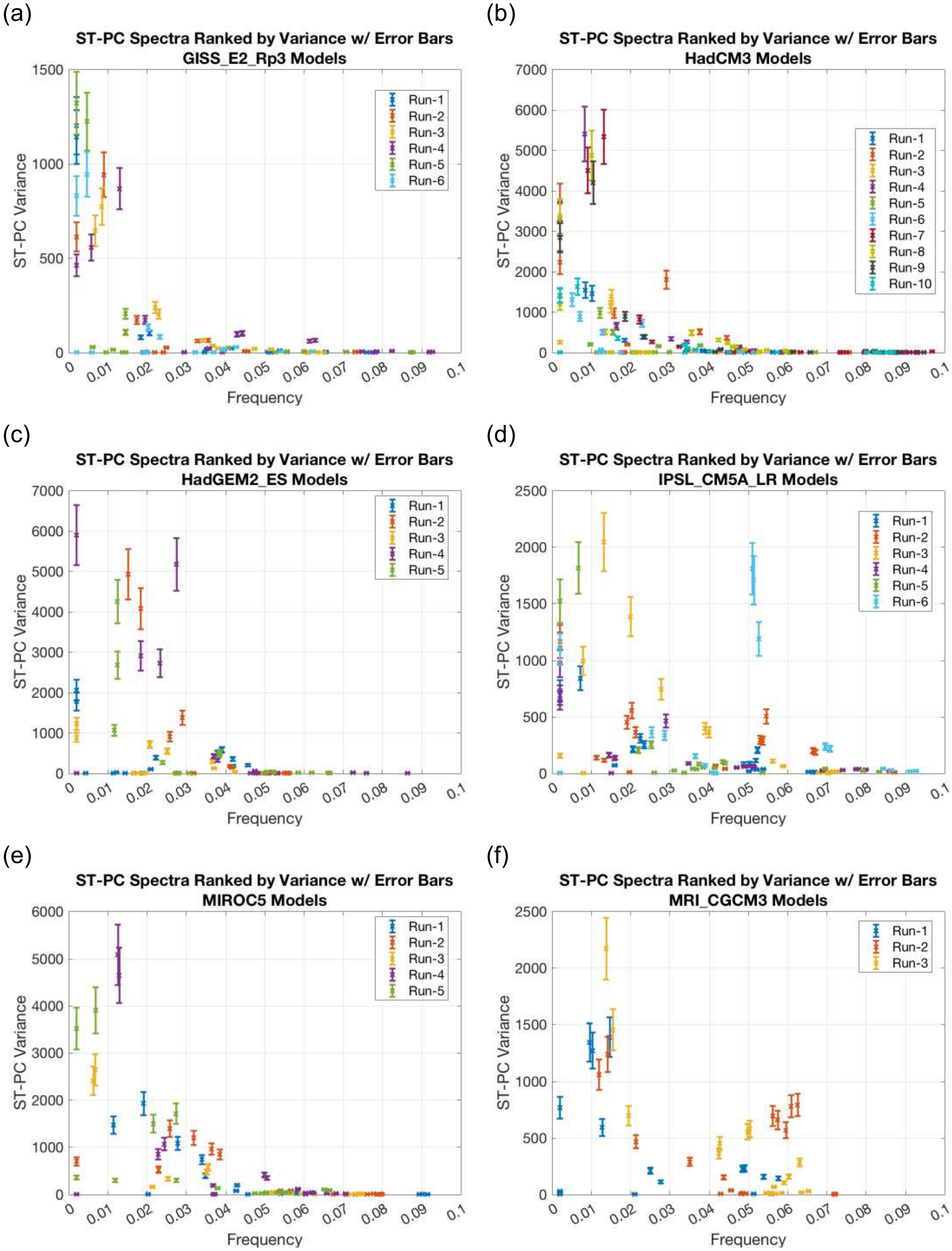


Figure 9: (a-f) Analogous to figure 7 except this time for the last six models listed in table 1 (models 12-17).

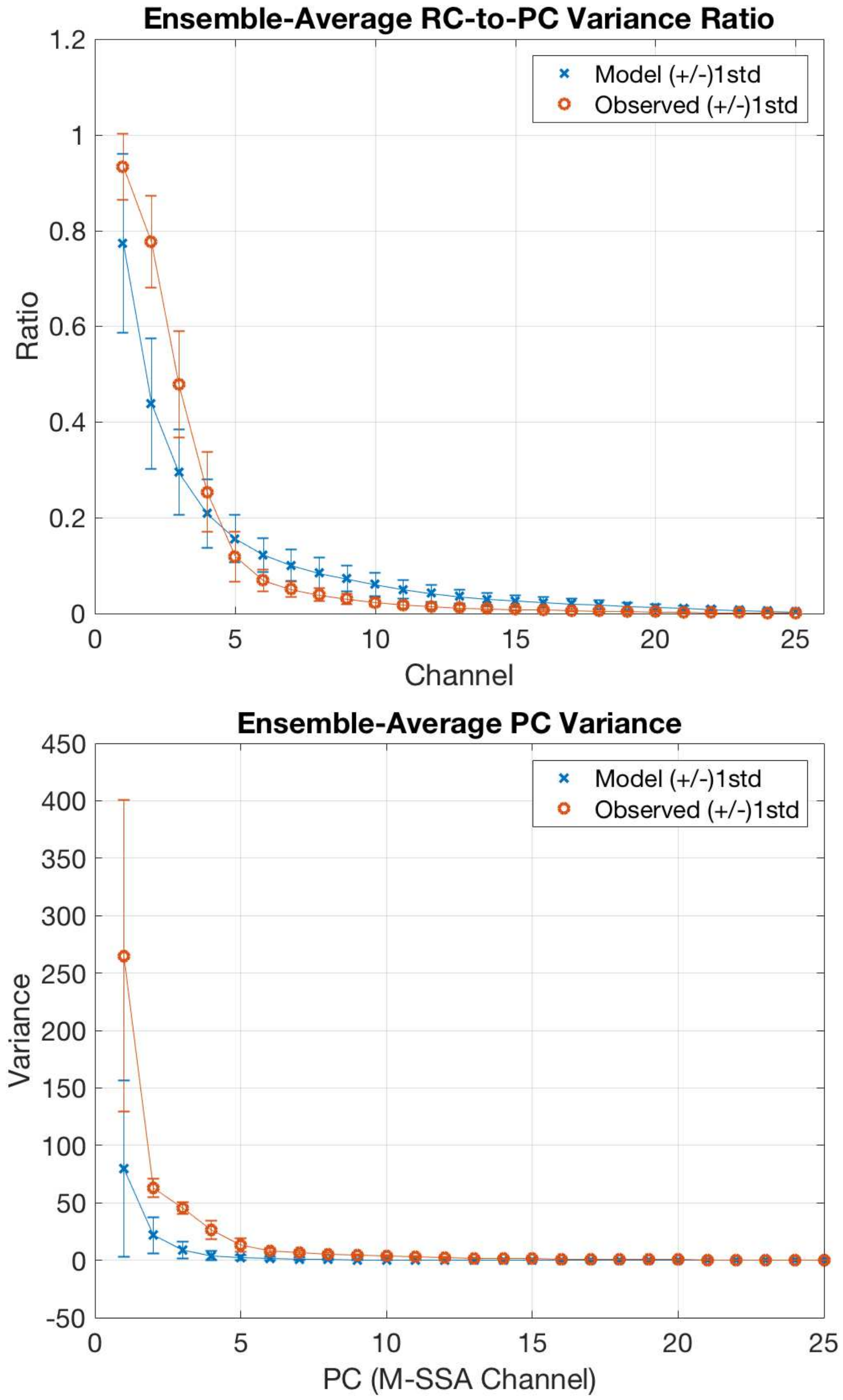


Figure 10: (Top) Ensemble-average variance ratio of Reconstructed components (RCs) of the leading M-SSA pair (modes 1 and 2) to original pre-filtered PCs in models (blue 'x') and observations (orange 'o'). This plot is a proxy for the "structure" of the variance within the two. **(Bottom)** Ensemble-average PC variances for models (blue 'x') and observations (orange 'o')

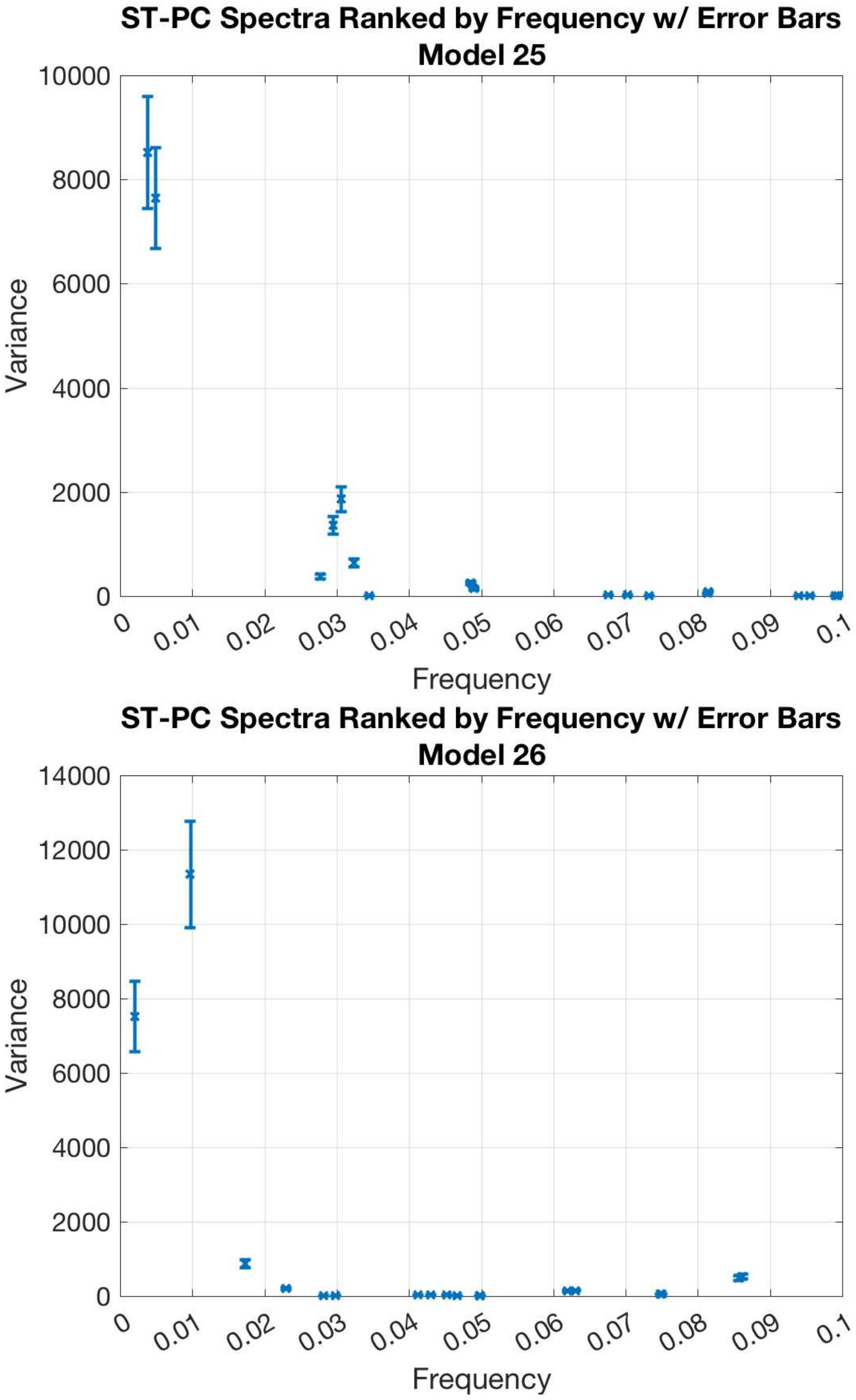


Figure 11: Frequency–variance M-SSA spectra for CSIRO model run 4 **(Top)** and run 5 **(Bottom)**. Both of these model runs fall within category 2 ("secular").

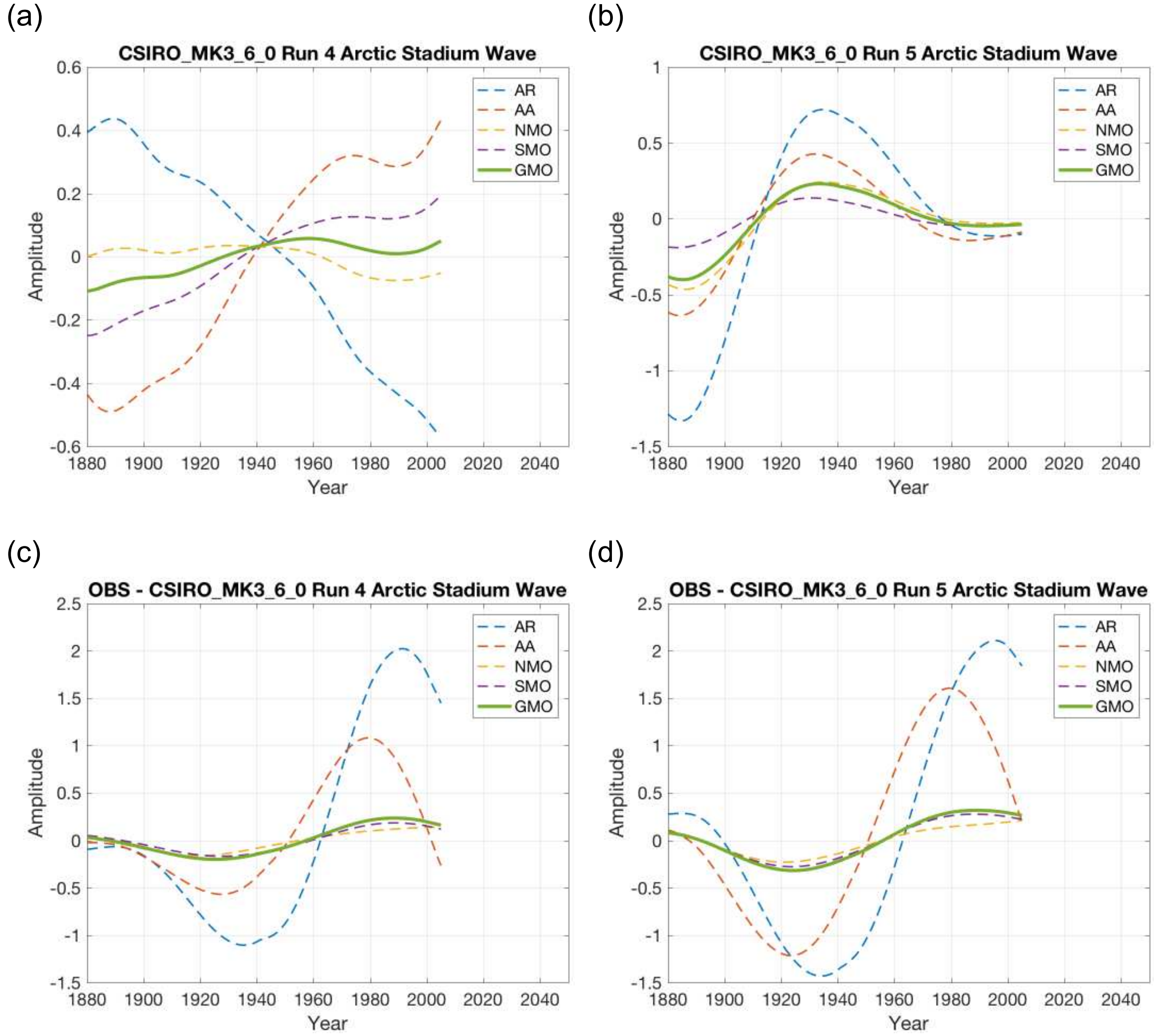


Figure 12: Regional index plots for CSIRO model run 4 **(a)** and run 5 **(b)**. Both of these model runs fall within category 2 ("secular") and show higher frequency oscillatory-signals superimposed on secular trends. The trend is most evident in run 4. The analogous figures, except for observations, are shown below them in **(c)** and **(d)**, respectively.

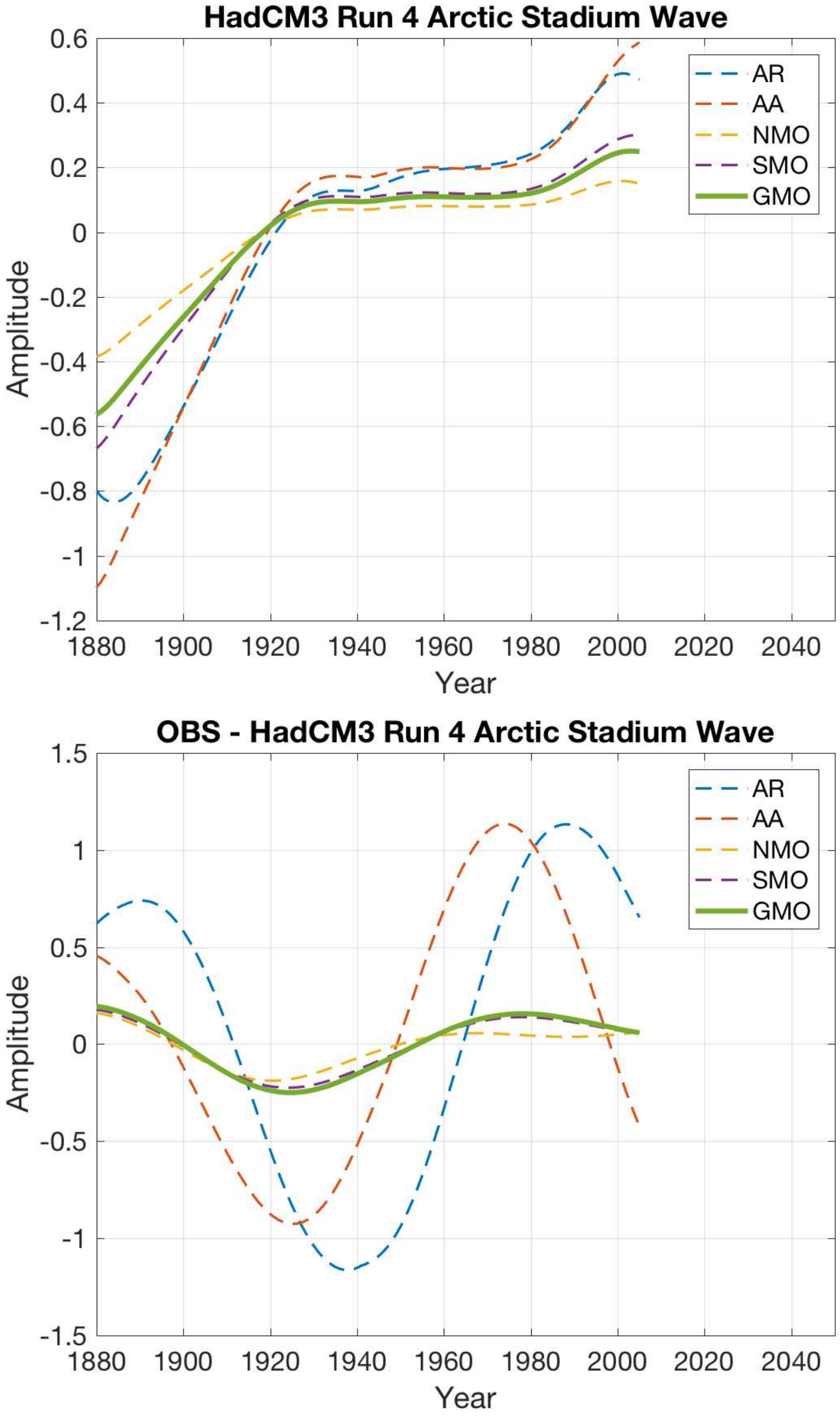


Figure 13: Analogous to the columns of Figure 12, except for HadCM3 Run 4.

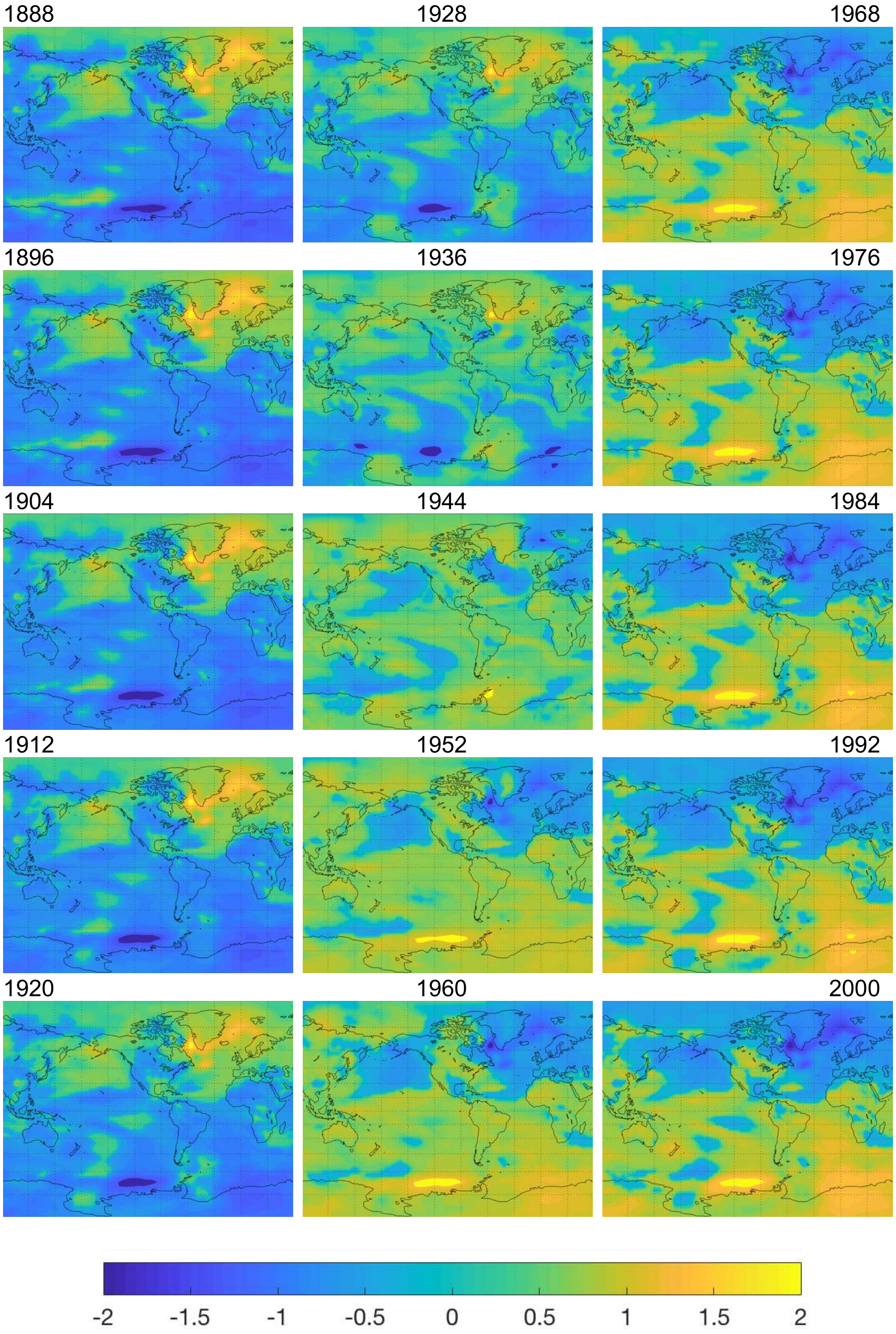


Figure 14: “Movie” of global CSIRO-MK3-6-0 Run 4 anomaly reconstructions from the leading M-SSA pair. The time-step between frames is eight years (from 1888 to 2000), encompassing the bulk of the 1880 to 2005 data timeframe. In order to accentuate the evolution of polar anomalies, all values were raised to the (1/3)-power.

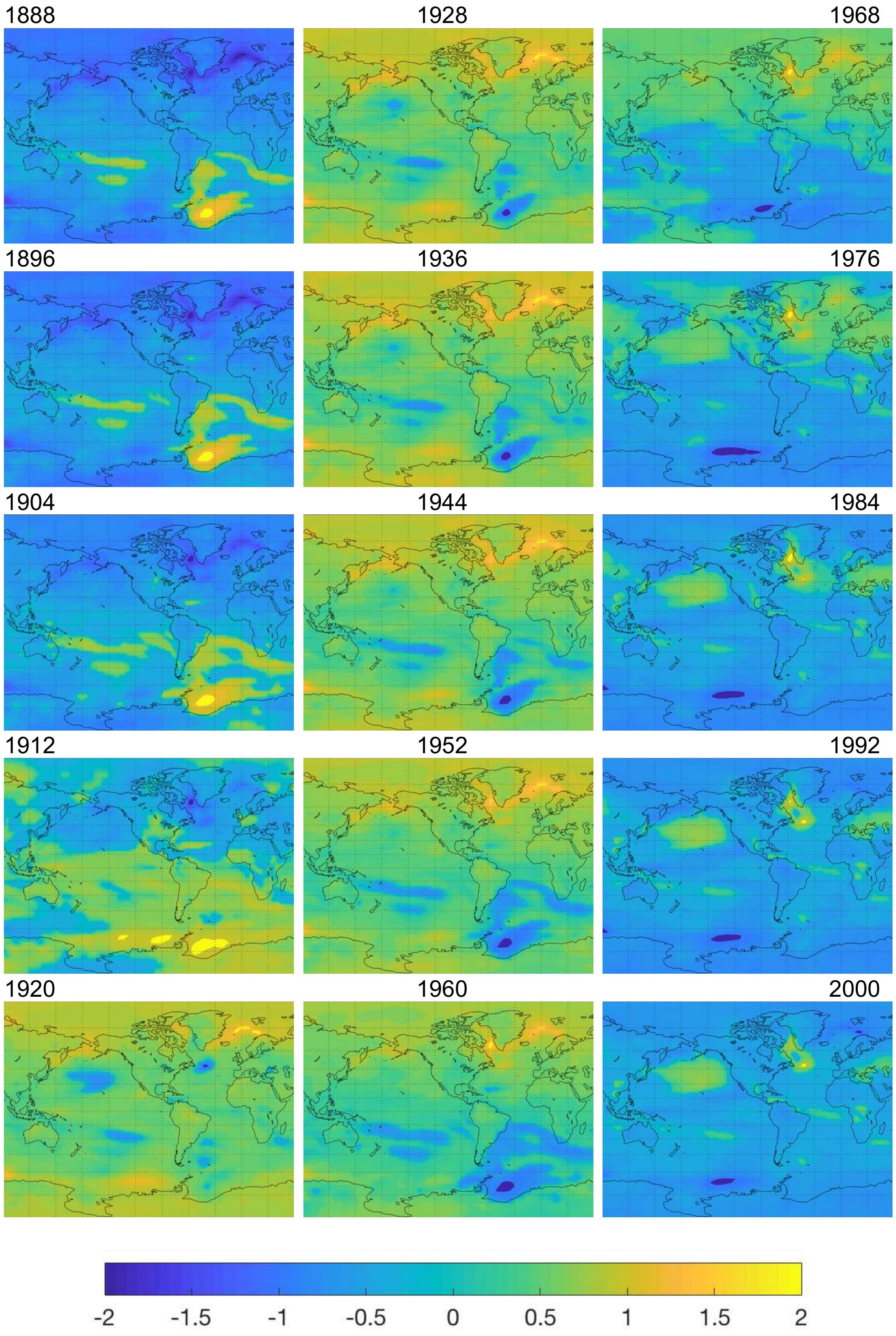


Figure 15: Analogous to figure 14 except for CSIRO-MK3-6-0 Run 5.

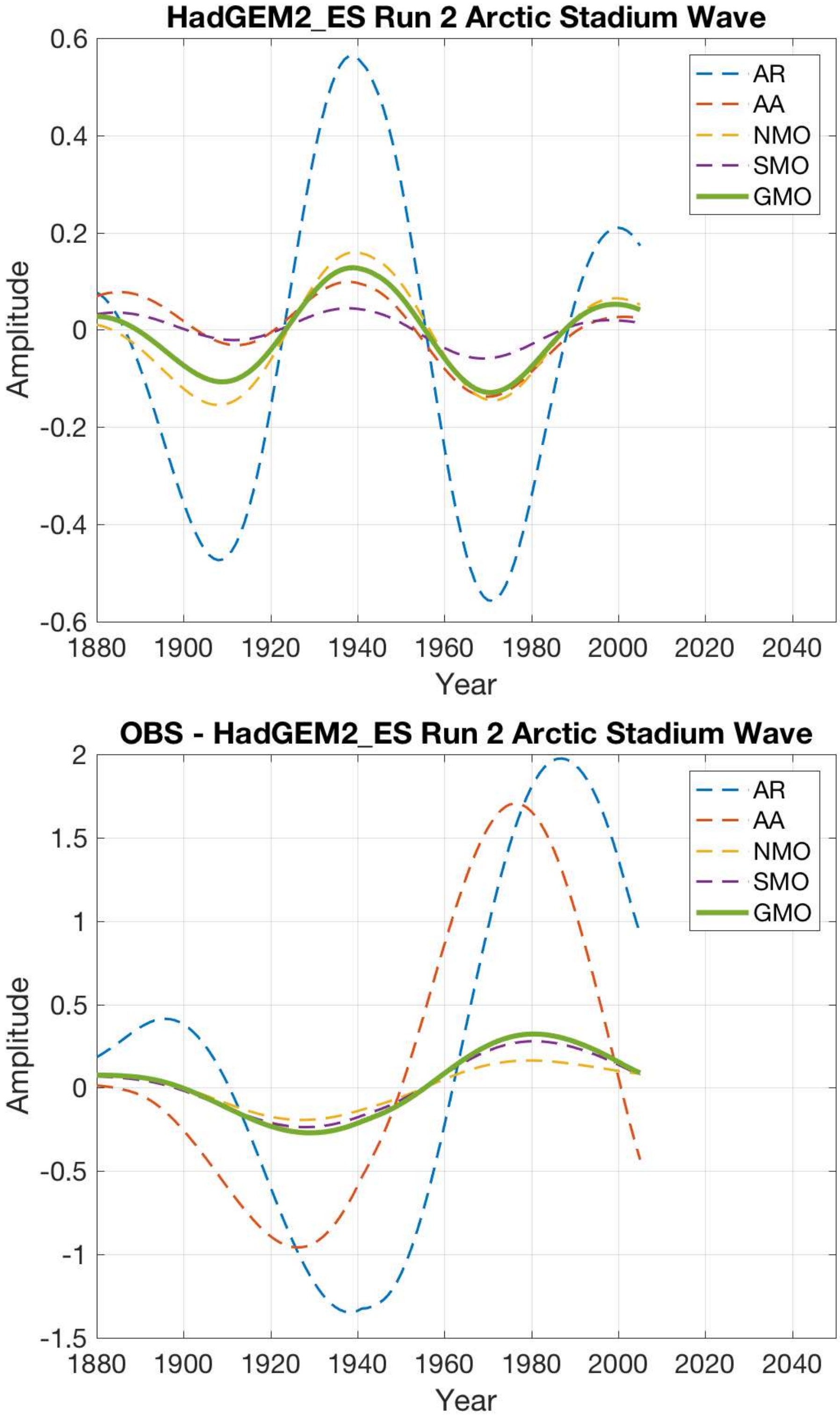


Figure 16: Analogous to Figure 13, except for HadGEM2-ES Run 2.

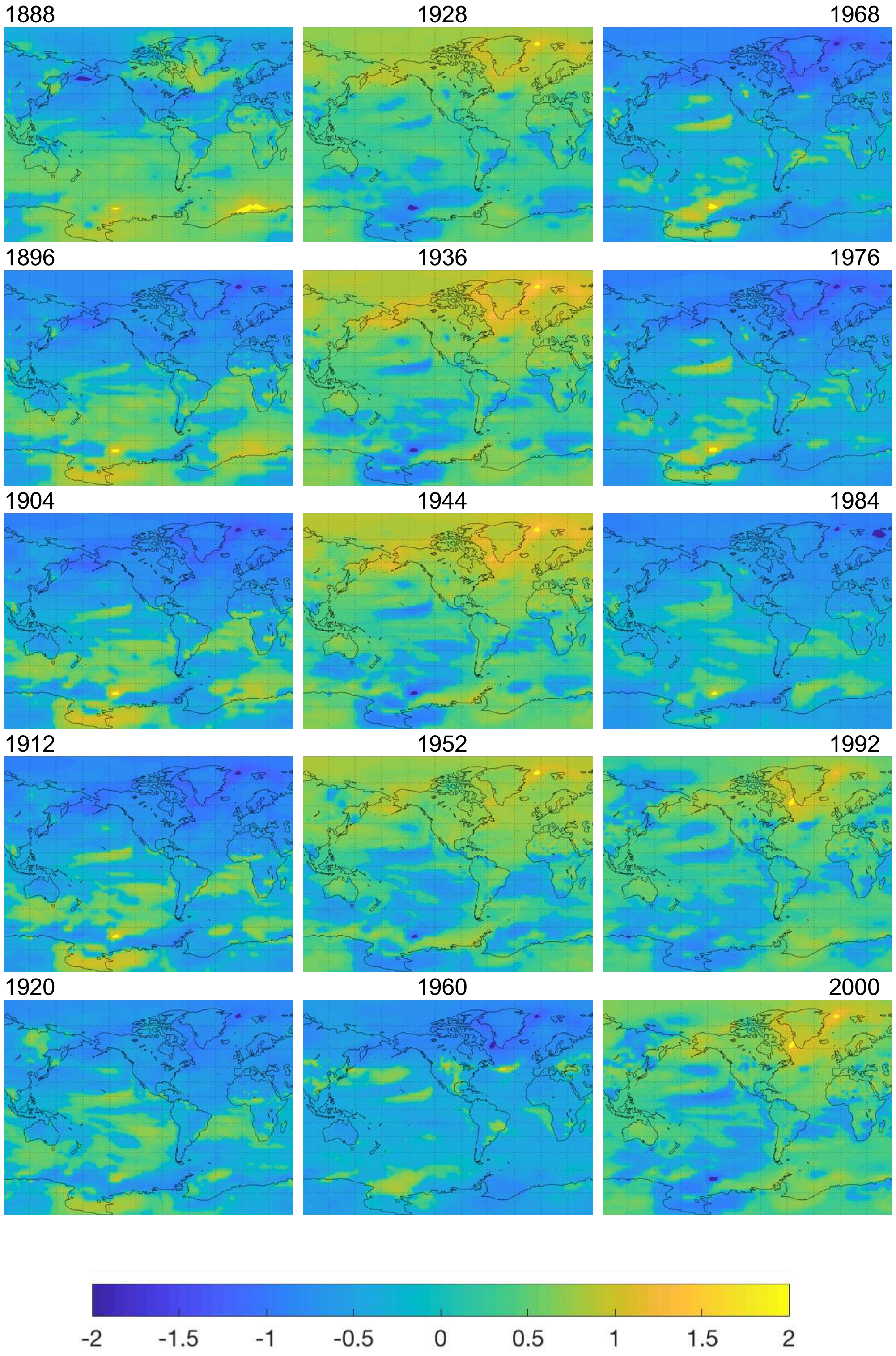


Figure 17: Analogous to figure 14 except for HadGEM2-ES Run 2.

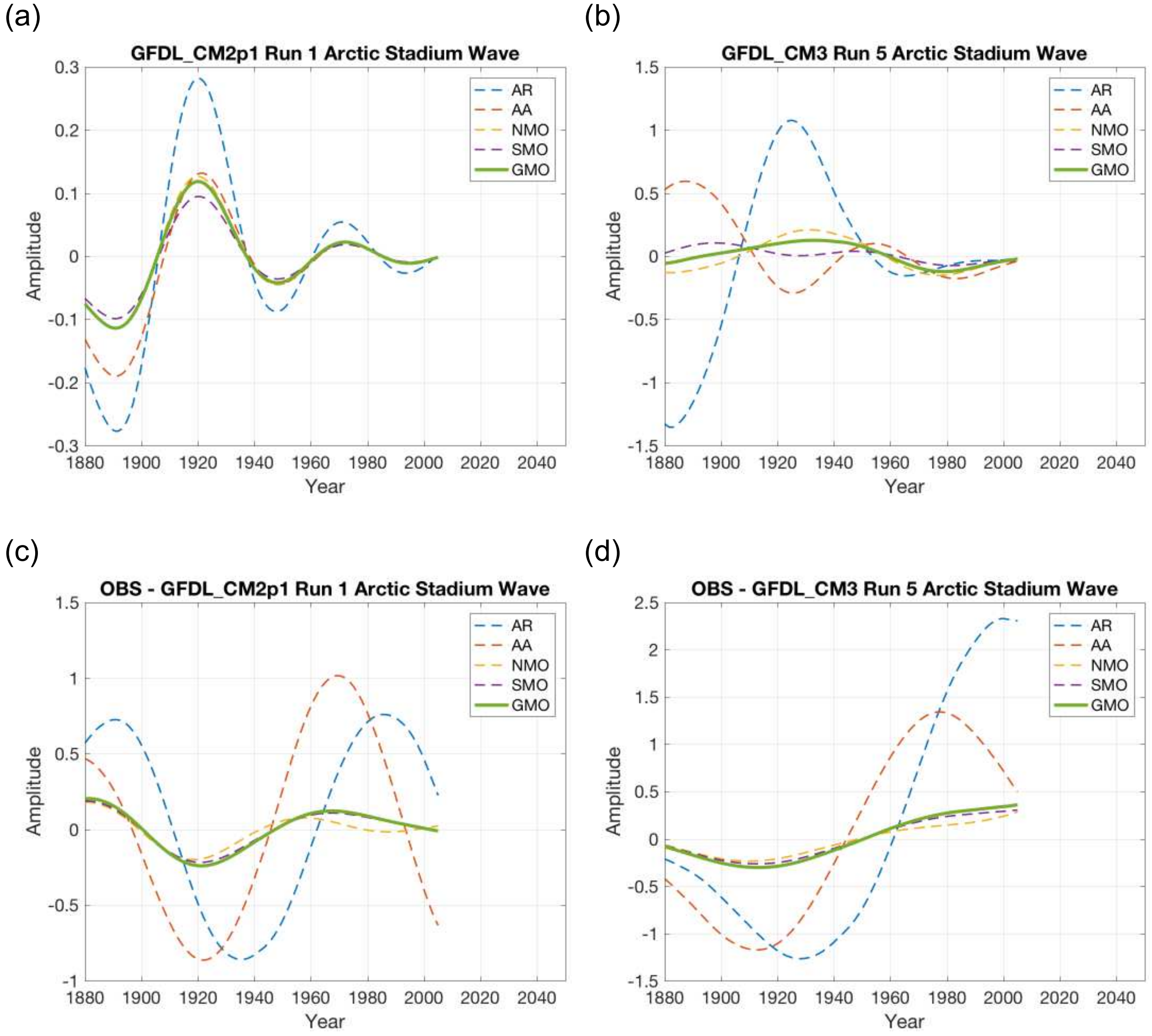


Figure 18: Analogous to Figure 12 except for GFDL-CM2.1 run 1 **(a)** and GFDL-CM2 run 5 **(b)**. Both of these model runs fall within category 3 (“non-pair”) and show in-phase and out-of-phase amplitude modulation in time, respectively. The analogous figures, except for observations, are shown below them in **(c)** and **(d)**, respectively.

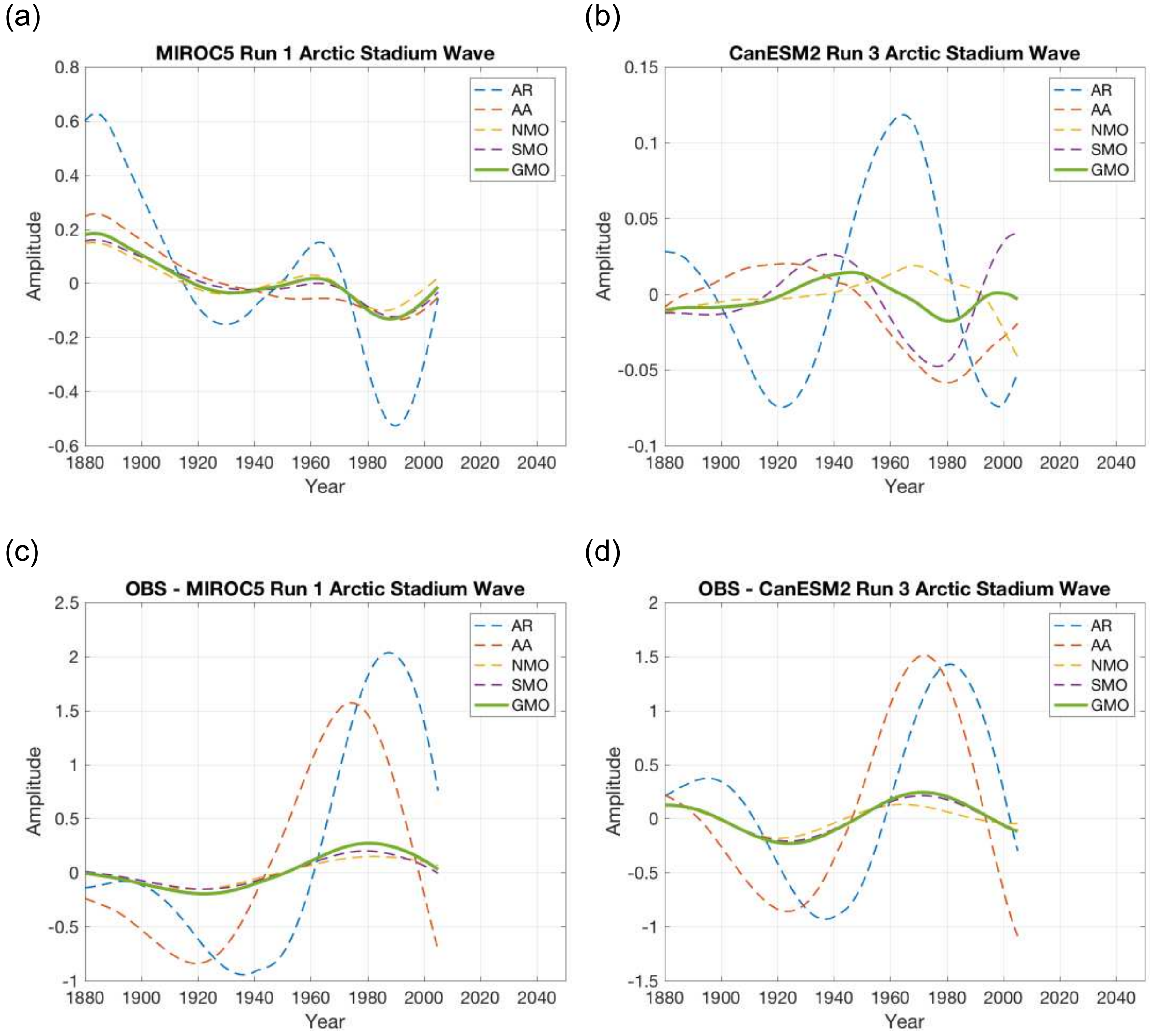


Figure 19: Analogous to Figure 12 except for MIROC5 run 1 **(a)** and CanESM2 run 3 **(b)**. Both of these model runs fall within category 3 (“non-pair”) a linear trend punctuated by oscillations and a mixture of two-or-more oscillatory signals, respectively. The analogous figures, except for observations, are shown below them in **(c)** and **(d)**, respectively.

Bibliography

Arakawa, A., 2004. The Cumulus Parameterization Problem: Past, Present, and Future. *Journal of Climate*, 17(13), pp.2493–2525.

Bellomo, K. et al., 2018. Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dynamics*, 50(9-10), pp.3687–3698.

Booth, B.B.B. et al., 2012. Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, 484(7393), pp.228–32.

Broomhead, D.S. & King, G.P., 1986. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2-3), pp.217–236.

Brown, P.T. et al., 2016. The necessity of cloud feedback for a basin-scale Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, 43(8), pp.3955–3963.

Buckley, M.W. & Marshall, J., 2016. Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review. *Reviews of Geophysics*, 54(1), pp.5–63.

Cassou, C. et al., 2018. Decadal Climate Variability and Predictability: Challenges and Opportunities. *Bulletin of the American Meteorological Society*, 99(3), pp.479–490.

Chylek, P. et al., 2014. Imprint of the Atlantic multi-decadal oscillation and Pacific decadal oscillation on southwestern US climate: past, present, and future. *Climate Dynamics*, 43(1), pp.119–129.

Compo et al., 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654), pp.1–28.

Danabasoglu, G. et al., 2016. North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part II: Inter-annual to decadal variability. *Ocean Modelling*, 97(C), pp.65–90.

Deser, C., & Phillips, A., 2016. The role of the North Atlantic Oscillation in European climate projections. *CLIVAR Exchanges/PAGES Magazine*, joint issue, https://doi.org/10.22498/pages.25.1.2.

van Dijk, A.I.J.M. et al., 2013. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, 49(2), pp.1040–1057.

Evan, A.T. et al., 2013. The Modification of Sea Surface Temperature Anomaly Linear Damping Time Scales by Stratocumulus Clouds. *Journal of Climate*, 26(11), pp.3619–3630.

Frankcombe, L.M. et al., 2018. On the Choice of Ensemble Mean for Estimating the Forced Signal in the Presence of Internal Variability. *Journal of Climate*, 31(14), pp. 5681–5693.

Ghil, M. et al., 2002. Advanced Spectral Methods for Climatic Time Series. *Reviews of Geophysics*, 40(1), pp.3–1-3–41.

Hedemann, C. et al., 2017. The subtle origins of surface-warming hiatuses. *Nature Climate Change*, 7(5), pp.336–339.

Hurrell, J.W., 1995. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269(5224), pp.676–679.

Hurrell, J.W. & Deser, C., 2009. North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, 78(1), pp.28–41.

IPCC, 2018. Summary for Policymakers. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. *World Meteorological Organization, Geneva, Switzerland*, 32 pp.

Jeffrey, S. et al., 2013. Australia's CMIP5 submission sing the CSIRO-Mk3.6 model. *Australian Meteorological and Oceanographic Journal*, 63(1), pp.1–13.

Kay et al., 2015. 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 the American Meteorological Society*, 96(8), pp.1333–1349.

Knutson, T.R., Zhang, R. & Horowitz, L.W., 2016. Prospects for a prolonged slowdown in global warming in the early 21st century. *Nature Communications*, 7, p.13676.

Kravtsov, S., 2017. Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. *Geophysical Research Letters*, 44(11), pp.5749–5757.

Kravtsov, S. & Callicutt, D., 2017. On semi-empirical decomposition of multidecadal climate variability into forced and internally generated components. *International Journal of Climatology*, 37(12), pp.4417–4433.

Kravtsov, S. et al., 2014. Two contrasting views of multidecadal climate variability in the twentieth century. *Geophysical Research Letters*, 41(19), pp.6881–6888.

Kravtsov, S. et al., 2016. Empirical Modeling and Stochastic Simulation of Sea Level Pressure Variability. *Journal of Applied Meteorology and Climatology*, 55(5), pp.1197–1219.

Kravtsov, S., Grimm, C., Gu, S., 2018. Global-scale multidecadal variability missing in state-of-the-art climate models. *npj Climate and Atmospheric Science*, 1(1), https://doi.org/10.1038/s41612-018-0044-6.

Kravtsov, S., Kondrashov, D. & Ghil, M., 2005. Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability. *Journal of Climate*, 18(21), pp.4404–4424.

Loehle, C., 2018. The Epistemological status of General Circulation Models. *Climate Dynamics*, 50(3), pp.1719-1731.

Martin, E., Thorncroft, C. & Booth, B., 2014. The Multidecadal Atlantic SST-Sahel Rainfall Teleconnection in CMIP5 Simulations. *Journal of Climate*, 27(2), pp.784–806.

Monahan, A. et al., 2009. Empirical Orthogonal Functions: The Medium is the Message. *Journal of Climate*, 22(24), pp.6501–6514.

Moron, V., Vautard, R. & Ghil, M., 1998. Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics, Berlin, Germany*, 14(7-8), pp.545–569.

Newman, M. et al., 2016. The Pacific Decadal Oscillation, Revisited. *Journal of Climate*, 29(12), pp.4399–4427.

Park, S. & Bretherton, C., 2009. The University of Washington Shallow Convection and Moist Turbulence Schemes and Their Impact on Climate Simulations with the Community Atmosphere Model. *Journal of Climate*, 22(12), pp.3449–3469.

Penland, C., 1989. Random forcing and forecasting using principal oscillation pattern analysis. *Monthly Weather Review, Boston*, 117(10), pp.2165–2185.

Penland, C., 1996. A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, 98(2-4), pp.534–558.

Penland, C. & Sardeshmukh, P.D., 1995. The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, 8(8), pp.1999–2024.

Poli et al., 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11), pp.4083–4097.

Qasmi, S., Cassou, C. & Boé, J., 2017. Teleconnection Between Atlantic Multidecadal Variability and European Temperature: Diversity and Evaluation of the Coupled Model Intercomparison Project Phase 5 Models. *Geophysical Research Letters*, 44(21), pp.11,140–11,149.

Robson, J. et al., 2012. Causes of the Rapid Warming of the North Atlantic Ocean in the Mid-1990s. *Journal of Climate*, 25(12), pp.4116–4134.

Ruprich-Robert, Y. & Cassou, C., 2015. Combined influences of seasonal East Atlantic Pattern and North Atlantic Oscillation to excite Atlantic multidecadal variability in a climate model. *Climate Dynamics*, 44(1), pp.229–253.

Schubert, S.D. et al., 2004. On the cause of the 1930s Dust Bowl. *Science (New York, N.Y.)*, 303(5665), pp.1855–9.

Smith, D. M., et al., 2016. Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, 6(10), pp.936–940.

Steinman, B.A., Mann, M.E., & Miller, S.K., 2015. Atlantic and Pacific multidecadal oscillations and Northern Hemisphere temperatures. *Science*, 347(6225), pp. 999–991

Stevens, B. & Bony, S., 2013. Climate change. What are climate models missing? *Science (New York, N.Y.)*, 340(6136), pp.1053–4.

Swanson, K., Sugihara, G. & Tsonis, A., 2009. Long-term natural variability and 20th century climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), pp.16120–16123.

Swingedouw, D. et al., 2015. Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions. *Nature Communications*, 6(1), pp.6545.

Taylor, K., Stouffer, R. & Meehl, G., 2012. An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4), pp.485–498.

Vautard, R. & Ghil, M. 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomena*, 35(3), pp.395-424.

Vautard, R., Yiou, P., & Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1-4), pp.95-126.

Wang, C. et al., 2012. Multidecadal Covariability of North Atlantic Sea Surface Temperature, African Dust, Sahel Rainfall, and Atlantic Hurricanes. *Journal of Climate*, 25(15), pp.5404–5415.

Wyatt, M. & Peters, G., 2012. A secularly varying hemispheric climate-signal propagation previously detected in instrumental and proxy data not detected in CMIP3 data base. *SpringerPlus*, 1(1), pp.1–25.

Yan, X.H. et al., 2016. The global warming hiatus: Slowdown or redistribution? *Earth's Future*, 4(11), pp.472–482.

Yeager, S. & Robson, G., 2017. Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability. *Current Climate Change Reports*, 3(2), pp.112–127.

Yuan, T. et al., 2016. Positive low cloud and dust feedbacks amplify tropical North Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, 43(3), pp.1349–1356.

Zhang, G.J. & Mcfarlane, N.A., 1995. Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean*, 33(3), pp.407–446.

Zhang, R. et al., 2013. Have Aerosols Caused the Observed Atlantic Multidecadal Variability? *Journal of the Atmospheric Sciences*, 70(4), pp.1135–1144.