Fingerprints of External Forcings on Sahel Rainfall: Aerosols, Greenhouse Gases, and Model-Observation Discrepancies

Kate Marvel[1], Michela Biasutti[2], and Céline Bonfils[3]

[1]NASA GISS/Columbia University
[2]Lamont-Doherty Earth Observatory
[3]LLNL

November 21, 2022

Abstract

Over the 20th and 21st centuries, both anthropogenic greenhouse gas increases and changes in anthropogenic aerosols have affected rainfall in the Sahel. Using multiple characteristics of Sahel precipitation, we construct a multivariate fingerprint that allows us to distinguish between the model-predicted responses to greenhouse gases and anthropogenic aerosols. Models project the emergence of a detectable signal of aerosol forcing in the middle of the twentieth century and a detectable signal of greenhouse gas forcing at the beginning of the twenty-first. However, the signals of both aerosol and greenhouse gas forcing in observations emerge earlier and are stronger than in the models, far stronger in the case of aerosols. The similarity between the response to aerosol forcing and the leading mode of internal variability makes it difficult to attribute this model-observation discrepancy to errors in the forcing, errors in the forced response, model inability to capture the amplitude of internal variability, or some combination of these. For greenhouse gases, however, the forced response is distinct from internal variability as estimated by models, and the observations are largely commensurate with the model projections.

Fingerprints of External Forcings on Sahel Rainfall: Aerosols, Greenhouse Gases, and Model-Observation Discrepancies

Kate Marvel[1,2], **Michela Biasutti**[3] **and Céline Bonfils**[4]

[1] NASA Goddard Institute for Space Studies, New York, NY USA

[2] Department of Applied Mathematics and Applied Physics, Columbia University, New York, NY USA

[3] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY

[4] Lawrence Livermore National Laboratory, Livermore, CA

E-mail: kate.marvel@nasa.gov

Abstract. Over the 20th and 21st centuries, both anthropogenic greenhouse gas increases and changes in anthropogenic aerosols have affected rainfall in the Sahel. Using multiple characteristics of Sahel precipitation, we construct a multivariate fingerprint that allows us to distinguish between the model-predicted responses to greenhouse gases and anthropogenic aerosols. Models project the emergence of a detectable signal of aerosol forcing in the middle of the twentieth century and a detectable signal of greenhouse gas forcing at the beginning of the twenty-first. However, the signals of both aerosol and greenhouse gas forcing in observations emerge earlier and are stronger than in the models, far stronger in the case of aerosols. The similarity between the response to aerosol forcing and the leading mode of internal variability makes it difficult to attribute this model-observation discrepancy to errors in the forcing, errors in the forced response, model inability to capture the amplitude of internal variability, or some combination of these. For greenhouse gases, however, the forced response is distinct from internal variability as estimated by models, and

the observations are largely commensurate with the model projections.

Submitted to: *Environ. Res. Lett.*

1. Introduction

Precipitation in the Sahel, the semi-arid region just south of the Sahara desert, affects a large and rapidly growing population. The region, which extends from Senegal in the west to Sudan and Ethiopia in the east, experiences rainfall concentrated in a wet season that runs June-October, with the bulk of precipitation falling in August and September. Spatially, the average rainfall varies sharply with latitude, with much smaller zonal gradients. It is strongly affected by internal climate variability at multiple spatial and temporal scales. Sahel rainfall is affected by the location of the Intertropical Convergence Zone (ITCZ), with increases in rainfall when the ITCZ is shifted anomalously north and drought when it moves south[?]. The region is also affected by variability in the global oceans[?, ?]: for example, the warming of the tropical troposphere during an El Niño event can suppress regional convection by enhancing atmospheric stability, while warming in the Atlantic or the Mediterranean can bring moisture to the region and strengthen the monsoon [?, ?, ?]. Despite these linkages to large-scale phenomena, however, aggregate rainfall in the Sahel results from short-lived weather systems on smaller time scales[?]. Understanding and simulating variability in Sahel rainfall therefore requires an integrated perspective of the drivers of this variability on multiple spatiotemporal scales[?, ?].

However, even against this noisy backdrop of internal variability the Sahel has experienced significant multidecadal precipitation changes over the 20th and 21st centuries attributable to external forcing[?, ?]. Idealized experiments with a single model attributed the pronounced 1950-1980 drying to external forcing[?], specifically anthropogenic aerosols. A study using CMIP3 and CMIP5 models and observations found that aerosol-induced cooling over the North Atlantic forces the ITCZ south, displacing the Sahel rain band[?], although this shift is severely underestimated by climate models. The severest recent drought in the region has been partially attributed to a combination of anthropogenic aerosol and volcanic forcing, notably the 1982 eruption of El Chichón[?].

In the wake of clean-air legislation passed in North America and western Europe, anthropogenic sulphate aerosol emissions fell and anthropogenic aerosol forcing decreased[?, ?]. In the 1990s, Sahel precipitation began to recover[?, ?]. However, the decrease in aerosol emissions has not been accompanied by a concurrent decrease in greenhouse gas emissions, which have continued to rise. To interpret the most recent trends, and to provide reliable projections of future rainfall, it is therefore crucial to disentangle the role of internal variability and multiple external forcings. If, for example, recent positive trends in Sahel rainfall result from a decrease in North America and western European aerosol [?], then we should not expect them to continue throughout the 21st century. If, however, they are attributable to greenhouse gas emissions[?], then we might expect future GHG emissions to accelerate existing trends, and plan accordingly.

Recent trends in mean Sahel precipitation suggest a recovery from the exceptionally dry conditions of the 1980s. But, while overall Sahel rainfall has increased, the spatiotemporal characteristics of the rainy season are also changing[?]. By 2007, precipitation in eastern regions of the Sahel had largely recovered, while the west was still considered in drought[?, ?]. Moreover, an increase in mean precipitation has also been accompanied by an intensification of individual rainfall events[?].

While it is difficult to formally attribute these observed changes to external forcing, they do appear to capture several aspects of the expected regional precipitation response to greenhouse gas forcing. An increase in mean and extreme rainfall intensity is a robust consequence of a warmer world, where increased latent heat flux from the surface is balanced by an increase in average precipitation[?] and the saturation water vapor pressure increases with temperature[?, ?]. Moreover, an east-west gradient in forced change is apparent in many climate models, possibly related to the zonal asymmetry in the Sahara Heat Low, an area of low pressure concentrated over the western Sahara. The low-level geostrophic flow into this local minimum advects dry subtropical air to the west and moist tropical air to the Eastern Sahel[?]. Warming strengthens this effect and contributes to drying in the West and wetting in the East: an asymmetry that should be exacerbated as greenhouse gases increase[?]. Finally, greenhouse gas forcing is expected to affect the seasonality of precipitation[?], with increases in rainfall largely confined to the late portion of the rainy season[?], when the barriers to convection in a more stable atmosphere are more easily overcome[?].

To what extent, then, do recent Sahel rainfall trends reflect the recovery from aerosol-induced drought versus the response to increasing greenhouse gas forcing? And is either of these responses distinguishable from internal variability? Previous work has attempted to separate the roles of these forcings by linking them to different SST warming patterns[?, ?]. Bonfils et al (in revision) employed a global-scale analysis of temperature, precipitation, and aridity to distinguish two modes of externally-forced fingerprints associated with greenhouse gas and aerosol forcing.

We build on this by employing the statistical techniques of multivariate detection and attribution to precipitation at a regional scale. Exploiting the coherent response to forcing across multiple variables, regions, or scales may both enhance the signal of anthropogenic forcing and decrease the noise, by rendering internal variability less likely to project by chance on the response to forcing. Here, we will use the characteristics of Sahel rainfall to create a multidimensional fingerprint that captures coherent aspects of expected forced change.

2. Methods

The simplest possible approach to detecting climate changes is to calculate the trends in regional or global mean variables and compare them to similar-length trends in unforced variability estimated by climate models[?, ?]. If such trends are deemed unusual in the context of model-estimated internal variability by some statistical test, they may be considered detectable. However, detecting the signatures of external forcings on regional climate is challenging for several reasons: internal variability may be considerable on regional scales, observational uncertainty may be large, and in some regions, high-quality observations do not exist over long timescales. Several authors [?, ?] have therefore advocated a process-based perspective that captures the specific spatial or seasonal aspects of the forced response in order to enhance the signal and decrease the noise. In detection and attribution research, the main goal is to separate the forced responses ("fingerprints") from the noise. In the literature, different statistical/numerical techniques exist to estimate these fingerprints (e.g., least-squares regression, optimal fingerprints with or without the need for EOF truncation and others methods [?, ?, ?]) . These may be trends, as discussed above, characteristic time series[?], or spatial patterns that capture the forced response[?]. Here, we will treat the "fingerprint" as a spatial pattern and define the fingerprint of a particular external forcing or collection of forcings as the leading empirical orthogonal function (EOF) of the average of model simulations run subject to those forcings[?]. Because the averaging process damps internal variability, the leading EOF generally explains a large proportion of the total variance[?, ?].

To track expected and observed spatiotemporal changes, we consider four indicators: two variables (monthly mean precipitation, hereafter PRMEAN) and the fraction of rainy days, defined as days where rainfall exceeds 1mm and hereafter referred to as R1) averaged over two regions (the central-eastern portion of the Sahel (east of the prime meridian) and the western Sahel (west of the prime meridian)). These quantities are calculated for CMIP5 historical simulations beginning in 1901. Because these historical simulations end in 2005, we extend them to the year 2100 by splicing with the corresponding RCP8.5 simulation beginning in 2006; we will refer to these extended simulations as H85. A list of all model simulations that provided relevant data for the historical and RCP8.5 simulations is provided in Table B1. Where multiple ensemble members are present, we calculate the multi-model average by first averaging over ensembles and then over models.

To ensure all variables carry the same units, we create z-scores $Z_X(t) = X(t)/\sigma_C^X$ by normalizing each variable $X(t)$ by a measure of noise σ_C^X. This is obtained by calculating monthly anomalies in X in the first 200 years of every pre-industrial control simulation, concatenating the resulting time series, and taking the standard deviation of the concatenated values.

To calculate the fingerprint, we construct the state vector

$$\mathbf{Z} = [Z_{PRMEAN(east)}, Z_{PRMEAN(west)}, Z_{R1(east)}, Z_{R1(west)}]$$

and perform the singular value decomposition

$$\mathbf{Z} = \mathbf{U\Sigma V}^T$$

where $\mathbf{\Sigma}$ is a diagonal matrix whose elements represent the eigenvalues. The unitary matrix $\mathbf{U}$ represents the multivariate EOFs, while $\mathbf{V}$ contains the principal components. The fingerprint is then defined as the leading multivariate EOF[**?**, **?**] and has 48 dimensions: it reflects changes in four variables over the twelve months of the calendar year. This fingerprint captures the model-predicted response to external forcing over *multiple* aspects of the Sahel rainy season.

The leading EOF of the H85 simulations is nonstationary: the fingerprint of external forcing varies with time because the forcings themselves vary with time. Figure 1 shows the fingerprints calculated from the H85 simulations over the twentieth (a) and twenty-first (b) centuries. The twentieth century fingerprint (hereafter 20CEN) is characterized by a symmetric decrease in precipitation in the rainy season across the eastern and western Sahel, and by a commensurate decline in the monthly fraction of rainy days. The twenty-first century fingerprint (21CEN), by contrast, is characterized by strong seasonality and spatial differences (Figure 1b), as has been noted previously[**?**]. In the eastern Sahel, rainfall increases throughout the rainy season. In the west, however, the pattern of change is characterized by a decrease in precipitation early in the rainy season and a smaller increase towards its end. The fingerprint is also characterized by changes in rainfall frequency, as measured by the total number of rainy days. The western Sahel experiences a decrease in the proportion of rainy days throughout the spring, summer, and fall, while the Eastern Sahel experiences decreases in the proportion of rainy days that are larger (and, in July, of opposite sign) than the changes in precipitation. The 21CEN fingerprint is nearly identical to the leading EOF calculated from the full 1901-2100 time period (Figure A1), while the 20CEN fingerprint strongly resembles the second EOF.

Here, we will argue that the 20CEN fingerprint largely captures the multi-model mean regional response to aerosols, while 21CEN captures the regional response to greenhouse gases. Aerosol forcing is believed to primarily affect Sahel precipitation remotely[**?**], by cooling the North Atlantic and forcing the ITCZ southward. This leads to a decrease in precipitation in the rainy season throughout the entire region:

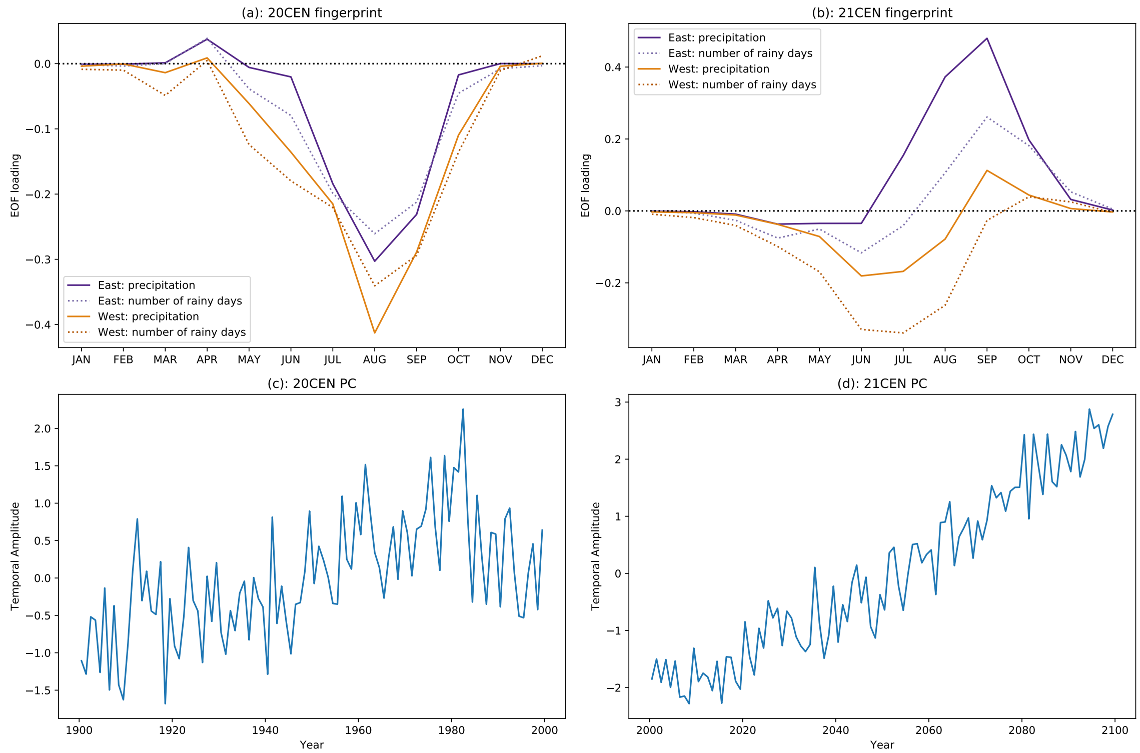

Figure 1. (a) Multi-variate fingerprint of forced changes in Sahel rainfall in the H85 simulations calculated over 1900-1999 (20CEN); solid lines depict total precipitation; dotted lines depict frequency of rainy days; purple is for Central and Eastern Sahel (east of the prime meridian) and orange is for the Western Sahel. (b) same as in (a), but over the period 2000-2099 (21CEN). (c) The principal component associated with the 20CEN fingerprint. (d) as in (c), but for 21CEN.

the response captured in the 20CEN fingerprint. The associated principal component (Figure 1c) tracks the temporal evolution of aerosol forcing, increasing through much of the twentieth century before peaking around 1980 and then decreasing. Greenhouse gas forcing dominates the RCP8.5 scenario, and the principal component associated with the 21CEN fingerprint (Figure 1d) reflects the monotonic increase in greenhouse gas emissions over the twenty-first century. The fingerprint itself captures many of the theoretically-expected characteristics of the response to greenhouse gas forcing: the asymmetry between the eastern and western Sahel, the seasonal variations, and the decoupling of mean precipitation change and number of rainy days. While the choice of the year 2000 as the dividing point between these two fingerprints is somewhat arbitrary, it does capture the fact that Western European and North Atlantic aerosols peaked and declined over the first period, and that the subsequent period is largely dominated by increasing greenhouse gas emission projected in RCP8.5. Other reasonable choices for the boundary between the aerosol-dominated period (for example, 2005, when historical simulations end, or 1990, slightly after the predicted peak in aerosol emissions) yield similar results.

The historical and RCP8.5 simulations are not forced by a single forcing agent, but by changes in anthropogenic (aerosols and greenhouse gases, but also ozone depletion and land-use changes) and natural (orbital changes, solar variability, and volcanic eruptions) forcings. It is therefore desirable to isolate the response to a single forcing by performing targeted simulations. Indeed, the CMIP5 archive contains some single-forcing simulations, namely those in which CO2 is increased at 1% per year (1pctCO2) and the subset of historicalMisc simulations run with only aerosol forcing. But there is a paucity of data in these simulations compared to the historical and RCP8.5 archives. Fewer models provided daily data (required to calculate R1) for 1pctCO2, and only four modeling groups provided the necessary data for the aerosol-only runs. The aerosol-only and CO2-only fingerprints can be calculated from these ensembles of reduced size and are shown in Figure A2. (We note that greenhouse gas-only "historicalGHG" simulations are also available in CMIP5, but we here use 1pctCO2 runs because more simulations of this type are available, and because the signal of greenhouse gas forcing is stronger in 1pctCO2 runs, resulting in a clearer fingerprint less contaminated by internal variability). They are qualitatively similar to the 20CEN and 21CEN fingerprints, respectively, but not exactly alike: the precipitation decreases in the aerosol-only fingerprint are confined to September and October, while the 1pctCO2 fingerprint shows more drying early in the rainy season in the eastern Sahel and does not show an increase late in the rainy season in the west. The differences between single-forcing and the H85 fingerprints, however, are artifacts of the reduced ensemble of models used. When 20CEN is re-calculated using only the models that provided aerosol-only simulations, the spatial correlation between this reduced-ensemble fingerprint and the aerosol-only fingerprint exceeds 0.95. When 21CEN is re-calculated using only the models that provided 1pctCO2 simulations, the spatial correlation between it and the 1pctCO2 fingerprint is 0.82. In order to utilize as many models as possible, here we will rely on the 20CEN fingerprint to approximate the CMIP5 multi-model mean response to aerosols, and on 21CEN to approximate the response to greenhouse gases.

The sensitivity of both fingerprints to the ensemble of models used indicates considerable uncertainty in the model responses to external forcings; this is reinforced by the comparatively small percentage of variance explained by CEN20 (27% of the variance in the 1900-1999 H85 multi-model mean) and CEN21 (50% of variance in the 2000-2099 H85 multi-model mean). Because the averaging process damps internal variability, the leading EOF of the multi-model average generally explains a large proportion of the variance- so why do CEN20 and CEN21 explain so little? First, the historical simulations are also forced by GHG and natural forcings, including volcanic eruptions that are intermittently quite large. These forcings have a response on Sahel rainfall that is non-negligible and not necessarily captured by the leading EOF. Similarly, while dominated, especially at the end of the 21st century, by greenhouse gases, the RCP simulations also include a reduction of anthropogenic aerosol forcing. Second, there is also considerable model disagreement in the response to single forcings: the aerosol-only and 1pctCO2 fingerprints explain only 22% and 36% of the multi-model average variance in the multi-model average of these ensembles, respectively. While this may be due to the smaller sample size of simulation and a less-clear separation between signal and noise, it is well-established that model uncertainties in forced responses may arise from biases in model climatology, particularly in the location of the major features of the circulation [?, ?], from differences in model dynamics[?, ?], uncertainty in the aerosol forcing itself[?], or differences in the model representation of aerosol direct and indirect effects[?].

The multivariate 20CEN and 21CEN fingerprints have the advantage of being nearly orthogonal to one another- the spatial correlation between the two is 0.1. (This can also be seen in the two leading EOFs of the total 1900-2099 H85 simulation, which resemble 21CEN and 20CEN, respectively, and are orthogonal by construction, Figure A1). This property means that, at least in models, it is possible to distinguish between the response to aerosol forcing (particularly as precipitation amounts recover from aerosol-induced declines) and the response to GHG forcing, as the leading response to one forcing will not strongly project on the fingerprint of the other.

The regional precipitation response to external forcing occurs against a backdrop of natural internal variability: climate "noise". Because we have no recent observations of unforced climate, and because paleoclimate proxies represent a climate forced by pre-industrial anthropogenic and natural forcings, we must rely on climate model pre-industrial control simulations (piControl) to characterize this variability[?]. We therefore calculate R1 and PRMEAN for the east and west Sahel in CMIP5 preindustrial control simulations, compute the anomalies, and concatenate the resulting time series. To prevent our results being dominated by models that performed extremely long piControl simulations, here we use only the first 200 years of each simulation. Figure 2 shows the three model-predicted leading noise modes. The primary mode of internal variability, figure 2a, is likely associated with northward and southward shifts in the ITCZ and is characterized by decreases in precipitation and number of rainy days throughout the entire Sahel in the rainy season. The decrease in rainy days accompanies the decrease in precipitation, indicating no substantial change in the intensity of rainfall. This mode strongly resembles the aerosol-dominated 20CEN fingerprint (the two patterns have a correlation above R=0.87). The second mode distinguishes between the early and late season but is distinct from the GHG-dominated 21CEN fingerprint in that East and West regions vary together and R1 tracks PMEAN. These results have important implications for the detectability of forced signals: because the 20CEN fingerprint is degenerate with the leading noise mode, models indicate that it will be more difficult to distinguish between the response to aerosols and internal variability. The same difficulty ought not affect the GHG signal.

To review, thus far we have presented fingerprints of the Sahel rainfall response to aerosols and greenhouse gases. The two are distinct from one another, indicating that a multivariate approach may be able to distinguish between the response to different forcings. Additionally, the GHG-dominated fingerprint is distinct from the leading modes of internal variability: in models, at least, internal variability does not project strongly on the predicted response to GHG forcing. The same is not true for the aerosol-dominated 20CEN fingerprint: because the response to aerosols strongly resembles the leading mode of internal variability, the models indicate that an aerosol signal must be extremely strong or persistent in order to become detectable over the background of climate noise.

3. Results

3.1. Model-predicted emergence time

We begin the detection and attribution analysis with a preliminary analysis to ascertain when the models themselves predict emergence of these signals. Given a dataset

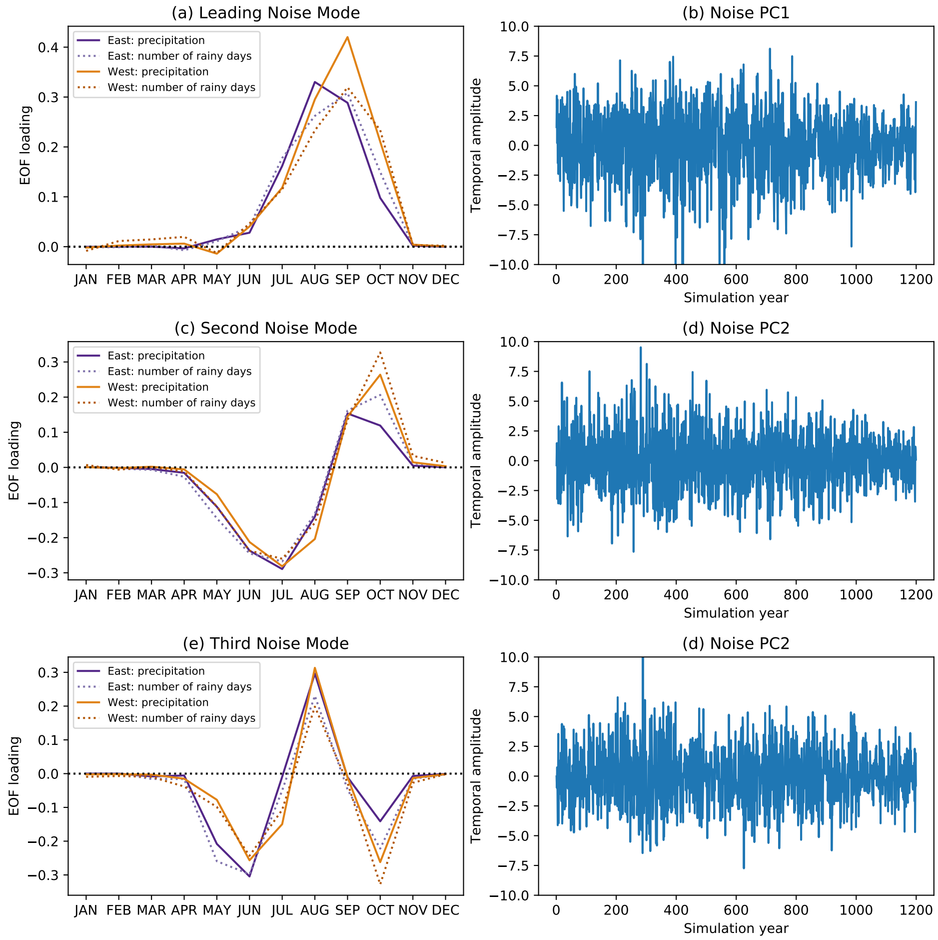

Figure 2. Leading multi-variate EOFs of Sahel rainfall natural variability as estimated by the CMIP5 pre-industrial control simulations. (a) First noise mode, explaining 17% of the total variance, and (b) the associated PC. (c) Second noise mode, explaining 12% of the total variance, and (d) the associated PC. (e) Third noise mode, explaining 11% of the total variance, and (f) the associated PC.. Line styles as in **??** and in legend.

beginning in 1901, when might we expect to see a detectable signature of these forcings? The model-predicted signal time of emergence is generally calculated using a standard framework employed in many previous detection and attribution studies[**?**, **?**]. For each model m, in each year t, we calculate monthly anomalies of PRMEAN and R1 in the east and west Sahel. We create z-scores by normalizing by σ_C^X and, as in calculating the fingerprint, create a state vector

$$\mathbf{Z^m(t)} = [Z^m_{PRMEAN(east)}(t), Z^m_{PRMEAN(west)}(t), Z^m_{R1(east)}(t), Z^m_{R1(west)}(t)].$$

We will define the *projection* as the dot product at every year t of this 48-element vector and the searched-for fingerprint. The resulting time series $P(t)$ measures the spatial covariance between the fingerprint and the observational or model data. If the fingerprint is increasingly present in the data, then $P(t)$ should trend upward. If the data is increasingly dissimilar to the fingerprint, then $P(t)$ should trend downward. Long-term changes in the projection time series therefore capture the resemblance between the searched-for fingerprint and the data, and we define the signal $S(L)$ as the L-length trend in $P(t)$, obtained by least-squares regression. This process reduces multidimensional data, varying across space, time, and multiple aspects of precipitation, to a single scalar signal.

Assessing the significance of such a signal requires an understanding of how internal variability– climate "noise"– could project onto the fingerprint due to chance alone. We therefore calculate z-scores from the CMIP5 preindustrial control simulations, normalizing, as before, with σ^X_C, the standard deviation of the concatenated anomalies. We project the resulting state vector $\mathbf{Z}_{control}$ onto the fingerprint to obtain a long time series $P_c(t)$. Because there is no *a priori* reason for internal variability to project positively or negatively on the fingerprint except by chance, the distribution of all possible L-length trends in this time series is here (and in general) well-approximated by a Gaussian with zero mean. We follow e.g. [**?**] in using the standard deviation of this distribution N(L) to define the noise. When the signal-to-noise ratio exceeds some pre-determined confidence interval, the signal is considered detectable. For example, when the signal-to-noise ratio exceeds 1.64, the observed signal is considered detectable at the 90% confidence level; in IPCC parlance it is "very unlikely" to be due to internal variability. The "time of emergence" is here defined as the year in which a signal, beginning in 1901, crosses this detectability threshold. If the detectable signal lies within the 90% confidence interval of simulations run subject to forcing, then it is "very likely" attributable to the forcing or collection of forcings in that simulation.

To determine the model-projected times of emergence, we calculate the projections $P_{20}(t)$ and $P_{21}(t)$ of each H85 model simulation onto the 20CEN (aerosol-dominated) and 21CEN (GHG-dominated) fingerprints, respectively. The multi-model mean projection $P_{20}(t)$ is shown as the thick pink line in Figure 3a; the 90% confidence interval of model $P_{20}(t)$ is shown as a pink shaded region. The multi-model mean is again calculated by averaging over ensemble members and then models; the model spread is calculated by projecting individual ensemble members onto the fingerprint. Figure 3b shows the multi-model mean projection $P_{21}(t)$ onto the 21CEN fingerprint (thick green line) and the 90% confidence interval determined by the H85 simulation projections. As indicated by the principal components in Figures 1c and 1d, on average H85 model simulations increasingly resemble the aerosol-dominated 20CEN fingerprint until roughly 1980, after which the projections trend negative. However, there is considerable uncertainty in the model projections onto this fingerprint throughout the 20th and 21st centuries; the 90% confidence interval, as determined by the spread in the model ensemble members, widens as the 21st century progresses, and we cannot say with confidence that the multi-model average projection onto 20CEN is positive or negative. There is considerably less ambiguity in the multi-model average projection onto the greenhouse gas-dominated 21CEN fingerprint, which remains positive throughout the 21st century.

Positive or negative trends in the projection time series can arise due to internal variability, and it is necessary to test such trends for detectability by calculating the

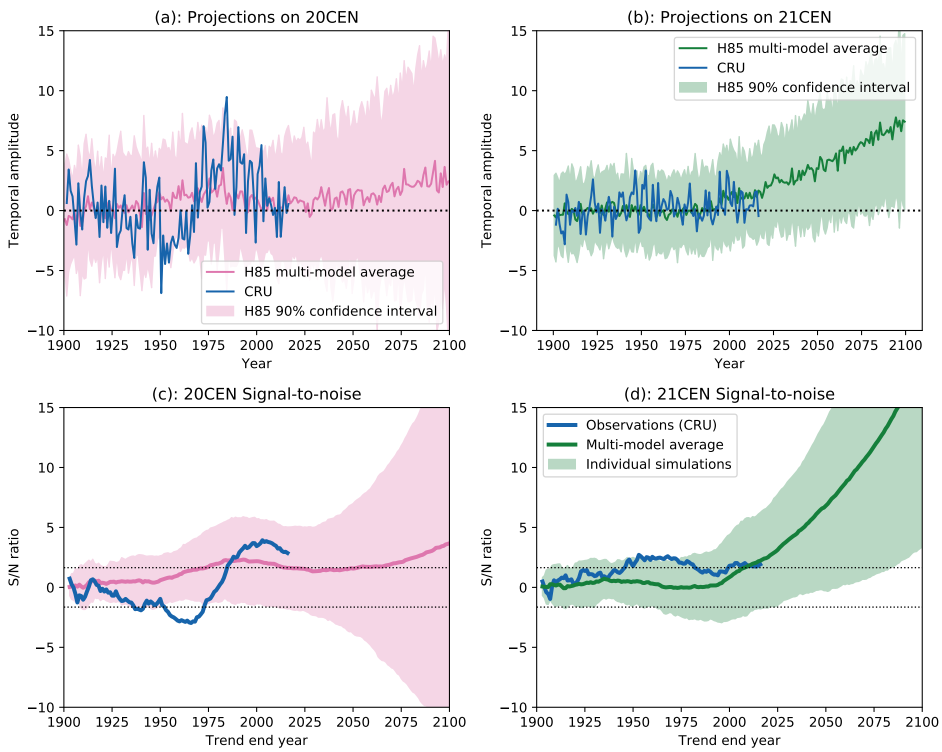

Figure 3. (a): Projection of observations and H85 simulations onto the 20CEN fingerprint. Thick pink line is for the mean across all H85 simulations and pink shading represents the 90% confidence interval as estimated from the ensemble spread. The blue line is the projection of observations on the 20CEN fingerprint. (b): same as (a), but for the 21CEN fingerprint. (c): Signal to noise ratio in forced trends. The signal is the magnitude of the trend in the 20CEN projection time series; the noise is the standard deviation of the distribution of trends in the pre-industrial control; trends start in 1900; the signal/noise ratio is plotted as a function of the trend end year. Pink line and shading for the H85 simulations and blue line for observations, as in (a). Values outside the dotted lines are detectable at the 90% level. (d) same as (c), but for the 21CEN fingerprint.

distribution of trends of the same length in the preindustrial control projection time series. To translate the projections in Figure 3a and 3b into signal-to-noise time series, we calculate the signal at time L for each H85 simulation, defined as the 1900-L trend in the projection, and divide by the corresponding noise term N(L)[**?**, **?**]. Figure 3c shows the signal-to-noise ratio for the aerosol-dominated 20CEN fingerprint. The aerosol-dominated 20CEN fingerprint (as predicted by the CMIP5 multi-model mean) becomes detectable at the "very likely" (90% confidence) level in 1982, peaks in 1993, then declines as aerosol forcing decreases. The models diverge in their 21st century behavior: some show a detectable resemblance to the 20CEN fingerprint, while some project that, under the RCP8.5 scenario, Sahel rainfall will become increasingly dissimilar. By contrast, the greenhouse gas-dominated 21CEN fingerprint first becomes detectable (in the multi-model mean) in 2017, and the signal-to-noise ratio increases with high confidence: the lower bound of the model ensemble-determined 90% confidence threshold crosses the detectability threshold before the end of the twenty-first century.

To illustrate the usefulness of the multivariate approach, we can also calculate model-projected times of emergence for individual variables (Figure A3). The GHG-dominated 21CEN fingerprint first becomes detectable in eastern Sahel rainfall in 2040, and in eastern Sahel rainy days in 2042, indicating that a multivariate fingerprint including these variables leads to earlier detection times than considering either individually. By contrast, the 21CEN fingerprint for mean rainfall and rainy days in the western Sahel becomes detectable in 1981 and 1983, respectively. But these early detection times result from the degeneracy between the aerosol-dominated (20CEN) and GHG-dominated (21CEN) effects on the Western Sahel. The models project a detectable signal of external forcing on the western Sahel rainy season emerging by the early 1980s, but are unable to distinguish between the responses to different external forcings. Only a process-based fingerprint that captures multiple aspects of Sahel rainfall change can distinguish between the responses to greenhouse gases and aerosols.

We note that these future projections are based on CMIP5's RCP8.5 simulations, which represent a plausible worst-case scenario but should not be construed as "business-as-usual". As more next-generation CMIP6 simulations begin to come online, modelers will be able to explore the consequences of more complex emission scenarios. In the RCP8.5 scenario used, global sulphur dioxide emissions sharply decline over the 21st century[**?**]. In other scenarios and in reality, the relative emissions of greenhouse gases and aerosols will change in the future in unpredictable ways. It is therefore imperative to stress that these future projections are dependent on a particular scenario, and may not represent the most likely future. As we will now show, there is another potential reason to question the results of the climate models: discrepancies between models and observations over the historical time period.

3.2. Detection and Attribution

To compare model output to observations, we use the gauge-based CRU TS dataset[**?**], which contains monthly time series of precipitation over Earth's land areas for 1901-2016 and uses the same station data to calculate the number of rain days R1. Its long extent and extensive validation make it useful for model-observation comparisons.

The CMIP5 models project the aerosol-dominated 20CEN fingerprint to be detectable in 1982, and the GHG-dominated 21CEN fingerprint in 2017. Despite substantial uncertainty in these signal emergence times, models indicate that detectable signals should be present in the observations. Are they?

Using the long-record CRU dataset, we calculate PRMEAN and R1 in the eastern and western Sahel between 1901-2016, normalize by σ_C^X, and calculate the projections $P_{20}(t)$ and onto the 20CEN fingerprint (blue line, Figure 3a). Over the 20th century, the observations appear increasingly dissimilar to the fingerprint until 1950, after which the resemblance sharply increases. Following the severe drought year of 1984, the fingerprint becomes less apparent in the observations. However, the observed trends in the projection $P_{20}(t)$ are far larger than in any model. This is reflected in the observed signal-to-noise ratio (blue line, figure 3c), which is far more variable than in any of the models. As indicated in Figure 3c, the 1900-1950 downward trend in the observations is large: over this time period the observations are increasingly dissimilar to the fingerprint. This trend is larger than in most piControl simulations. Post 1950, the observed $P_{20}(t)$ trends upward (Figure 3c). The observed signal-to-noise ratio begins to increase and exceeds the 90% detectability threshold in 1987. The 1901-1987 trend is formally detectable and attributable (ie, inconsistent with model-estimated noise but consistent with the forced model distribution), and the signal remains (formally) detectable through the present. This should not, however, obfuscate the fact that while models and observations may agree over this centennial time scale, there is a substantial disconnect between models and observations at shorter, multidecadal scales.

On these shorter time scales, there is clearly more variability in the projection of the observations onto 20CEN than in the model projections onto the same fingerprint: Figure 3(a) indicates a larger observed negative trend and more significant signal-to-noise ratio (Figure 3(c)) than in the CMIP5 models. This suggests that models underestimate multidecadal variability because they fail to capture the full spectrum of low-frequency internal variability present in the real world, a realistic response to aerosol forcing, or both. This analysis cannot distinguish between the two, since the response to aerosol forcing (Figure 1a) in these variables so closely resembles internal variability (Figure 2a). The large 1900-1950 negative trend, over a time period when aerosol forcing was small, compared to its later peak in 1950-1980, but increasing may constitute evidence for the hypothesis that models fail to capture multidecadal internal variability.

The blue line in Figure 3d shows the observed signal-to-noise ratio for the GHG-dominated fingerprint 21CEN. In the observations, the signal becomes detectable as early as 1940 before decreasing, finally re-emerging at the 90% detectable level in 2001. The 1900-1950 signal lies at the edge of the 90% confidence interval of the H85 models, indicating that, very likely, either models underestimate the greenhouse gas response, there exists a mode of internal variability not simulated by climate models that resembles the greenhouse gas fingerprint, or both.

The linear approach to signal detection is complicated by forcings that do not increase or decrease monotonically. For example, the 20CEN fingerprint is detectable in 1901-2016 observations and compatible with model trends in $P_{20}(t)$ over the same period, but this model-observation agreement masks the fact that, in the observations, a large positive 1950-1980 trend is preceded by a large negative 1900-1950 trend, rendering the aggregate 1901-2016 trend less positive and therefore compatible with the smaller model-predicted trends. We therefore perform the linear detection analysis on three periods: 1900-1950, where the observed trend is negative, 1950-1980, where it is positive, and 1980-2016. Figure 4 shows the resulting signal-to-noise ratios for both the 20CEN and 21CEN fingerprints. Model-predicted signals of 20CEN are shown as pink lines; the observed trends are shown as white circles. Over the 1950-1980 period, the observed 20CEN signal is larger than in either forced or unforced model simulations. However, the distribution of simulated forced trends is not distinguishable from the distribution of unforced trends arising from the model-estimated internal variability, muddling a clear attribution to increasing AA forcing. The 1980-2016 trend is again far more negative than in forced or unforced model simulations. This time, the distribution of simulated forced trends is distinguishable from the distribution of unforced trends, suggesting that a decline in the impact of aerosols after the Clean Air Act is operating in both observations and models. However, the model response underestimates the amplitude of the observed response.

The observed projections onto the greenhouse gas-dominated 21CEN fingerprint tell a different story. The observed signal of greenhouse gas forcing is detectable in the 1900-1950 period and compatible with the H85 model simulations. Both the decreasing

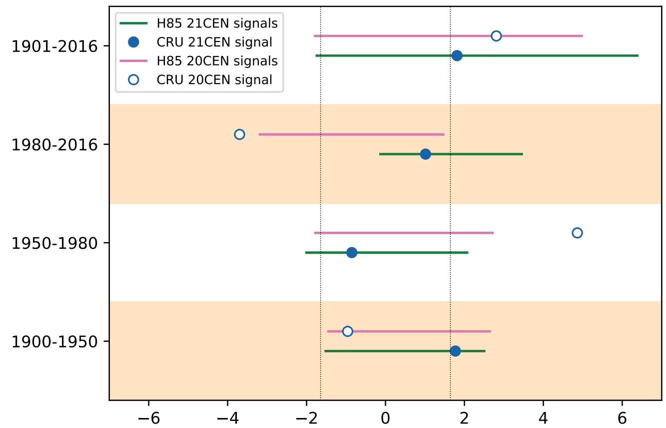

Figure 4. Simulated and observed trends in the projections on 20CEN (pink lines and open circles) and 21CEN (green lines and filled circles) fingerprints for the periods 1900-1950, 1950-1980, 1980-2016, and 1901-2016. Pink and green lines depict the 90% confidence interval determined by the spread in H85 simulations; circles depict observed values; the vertical dotted lines mark the 90% confidence intervals for the signal-to-noise ratio.

similarity between the observations and the 21CEN fingerprint over the 1950-1980 period and the subsequent increase in observed $P_{21}(t)$ from 1980-2016 do not exceed the detectability threshold, and are compatible with model-simulated internal variability. The longer 1900-2016 trend is just outside the 90% confidence interval for unforced variability and might therefore be considered detectable and attributable to greenhouse gases. The fact that the signal emerges in 1950 but subsequently wanes even as the forcing increases suggests that, in the real world, the effects of aerosol forcing or internal variability may mask the projection onto the greenhouse gas-dominated fingerprint. Because of this, and given the inability of models to capture the observed trends in projections onto the 20CEN fingerprint, we urge caution in interpreting these projections onto the 21CEN fingerprint.

4. Conclusions

Detecting and attributing changes in regional precipitation is challenging due to uncertainty in model responses to multiple forcing agents and the large amplitude of internal variability. Here, we have adopted a process-based approach to fingerprinting, exploiting coherent responses across multiple variables to distinguish the signals of different external forcings. In models, the seasonality and east-west gradient of change differ under greenhouse gas and aerosol forcing, resulting in multivariate fingerprints that are distinct from one another.

However, we show substantial differences in the modeled and observed projections onto the 20CEN fingerprint, a pattern we argue reflects the multi-model mean response to aerosol forcing. This is complicated by the fact that this aerosol-dominated fingerprint (Figure 1a) so closely resembles the leading noise mode (Figure 2a) in climate models. This means that, in addition to the usual explanations for an observed signal-to-noise ratio being higher than that in models (the observed signal is stronger than in models, or the model-estimated noise term is too small), it may also be the case that the degeneracy between the fingerprint representing the forced response and the pattern characteristic of the leading noise mode in the models, which delays the emergence of the signal, does not hold in the real world.

The projection on the 21CEN fingerprint increases somewhat faster than expected (at the 90% level) during the first half of the 20th century, and the mismatch is not easily interpreted as a bias in the model-simulated noise. This is because the model response to greenhouse gas forcing does not strongly resemble model-simulated climate noise modes. It is possible that this is due to errors in simulated internal variability, but this would suggest models fail to capture an important mode or modes of climate noise, not just their amplitude. It is also possible that climate models fail to capture the strength of the response to greenhouse forcing. However, the observed greenhouse gas-dominated 21CEN signal is detectable from 1901-1950, and while the observed trend is larger than most forced runs, it is still compatible with the forced distribution. The signal is subsequently lost at mid-century, likely due to a masking effect from the aerosols, and then reappears. In each case, it is compatible with the simulated trends.

Finally, while we show that in the limited-size ensemble of models that performed single-forcing simulations, the 20CEN fingerprint resembles the aerosol-only fingerprint and the 21CEN fingerprint resembles the CO_2-only fingerprint, it is important to note that historical simulations and observations contain the climate response to *multiple* external forcing factors, both natural and anthropogenic. It is useful to show as we do here, that multivariate fingerprints can distinguish between aerosol-dominated and greenhouse gas-dominated responses in models. However, multiple studies[**?**, **?**] have identified a signature of naturally forced change in Sahel precipitation over the twentieth century. Larger ensembles of single-forcing simulations are needed to more clearly identify the model responses to natural forcings and distinguish these from aerosol and greenhouse gas responses.

What is the way out of this impasse? The transition from CMIP3 to CMIP5 or, as shown by a preliminary analysis [**?**], to the CMIP6 generation of climate models has not solved these issues. Regional simulations that can explicitly simulate mesoscale systems and thus the intensity characteristics of Sahel rainfall are coming online, but their response to external forcing still depends on the boundary conditions simulated by coarser GCMs [**?**], so that reasons for doubt persist. Moreover, even at global scales, the spread in the CMIP6 climate model ensemble appears to be increasing[**?**]. Nevertheless, we can better understand the sources of model biases in Sahel rainfall variability if we make strategic use of different sets of multi-model simulations: in idealized configurations, of high-resolution atmosphere-only models forced by SST, in large-ensemble historical coupled simulations, and in those initialized for decadal predictions. Such a concerted effort will potentially reduce uncertainty in the regional response to aerosol forcing, in the history of internal modes of SST variability at decadal time scales, and in the role of convective-scale processes – both in the atmosphere and at the land surface – in shaping structural uncertainty in model responses. The many MIP ensembles now coming online as part of CMIP6 provide an opportunity to advance our knowledge of forced Sahel rainfall trends in the next decade.

5. Acknowledgements

K.M. and C. B. were supported by the US Department of Energy Biological and Environmental Research Grant DE-SC0014423. Work at LLNL was performed under the auspices of the US Department of Energy under contract DE-AC52-07NA2734

6. Data Availability

The data that support the findings of this study are openly available. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in the supplementary tables of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

A Supplementary Figures

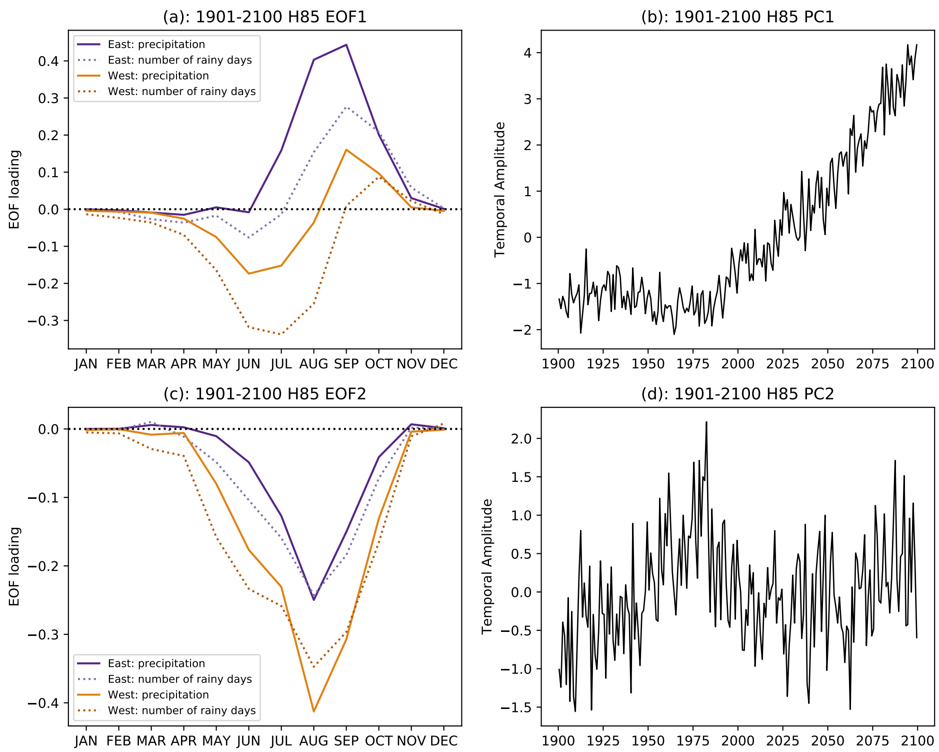

Figure 1: Fingerprints calculated using 1900-2100 H85 model simulations: (a) EOF1 and (b) the associated PC; (c) EOF2 and (d) the associated PC

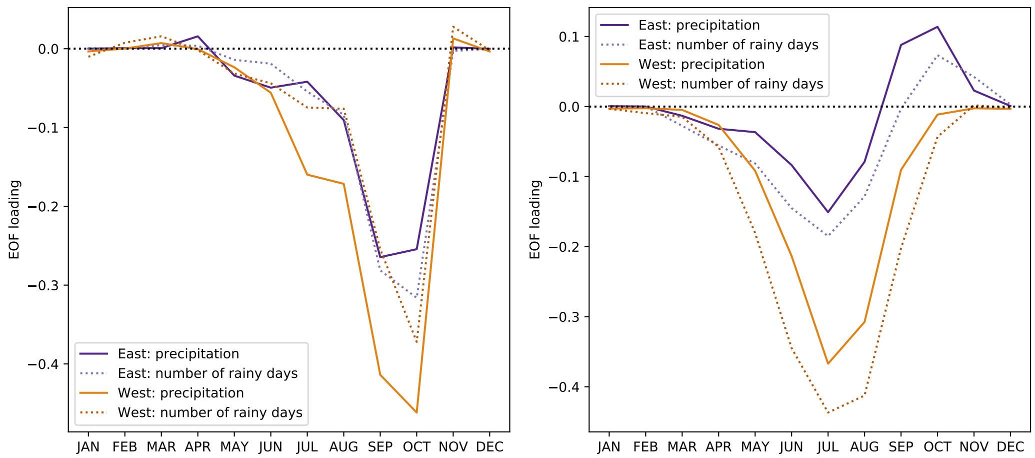

Figure 2: Fingerprints calculated using (a) CMIP5 aerosol-only "historicalMisc" simulations over the period 1900-2005 and (b) CMIP5 "1pctCO2" simulations

B Supplementary Tables

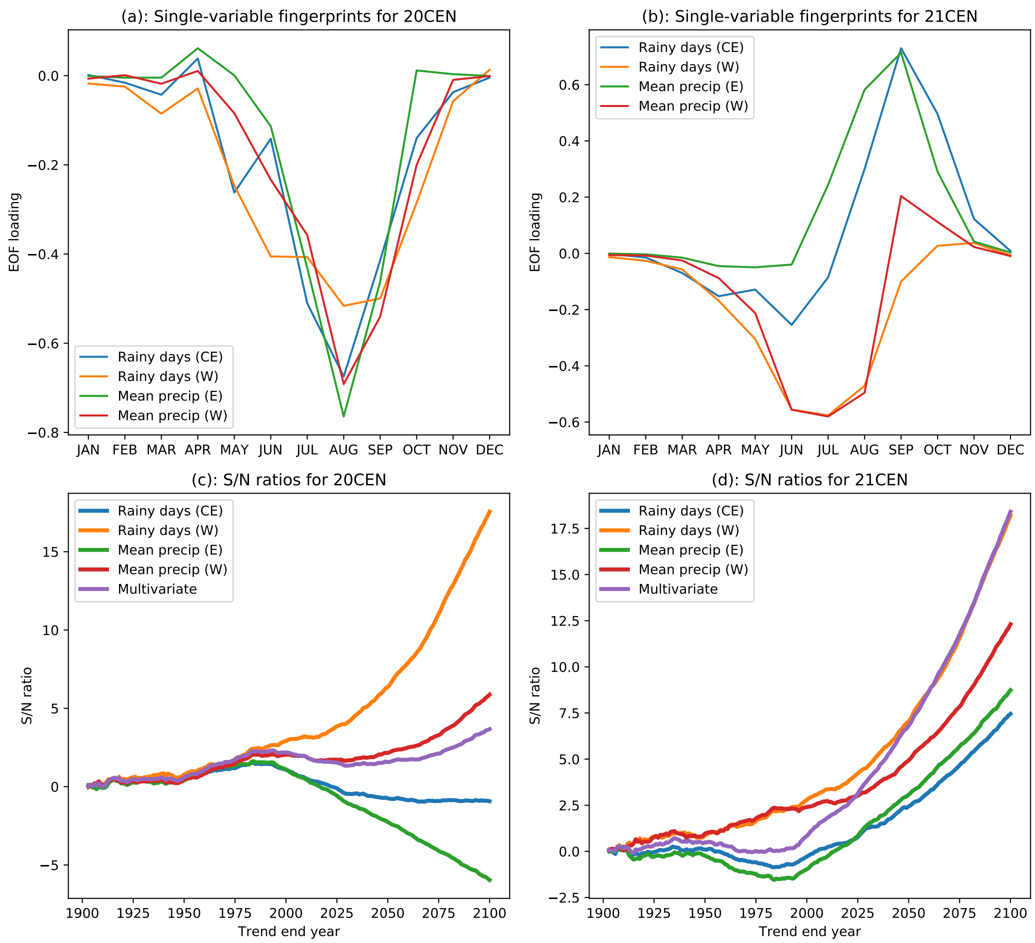

Figure 3: Single-variable fingerprints calculated from H85 simulations for (a) the 20th century and (b) the twenty-first century. We note the resemblance of these fingerprints to their counterparts in the multivariate fingerprints. (c): Multi-model average signal-to-noise ratio for the emergence of the 20CEN fingerprint in indiviual variables (d): Same as (c), but for the 21CEN fingerprint

Modeling group name	simulation ID
CCSM4	r1i1p1
CCSM4	r2i1p1
CESM1-CAM5	r1i1p1
CMCC-CM	r1i1p1
CMCC-CMS	r1i1p1
CNRM-CM5	r1i1p1
CSIRO-Mk3-6-0	r1i1p1
CanESM2	r1i1p1
CanESM2	r2i1p1
CanESM2	r3i1p1
CanESM2	r4i1p1
CanESM2	r5i1p1
GFDL-CM3	r1i1p1
HadGEM2-AO	r1i1p1
IPSL-CM5A-LR	r1i1p1
IPSL-CM5A-LR	r2i1p1
IPSL-CM5A-LR	r3i1p1
IPSL-CM5A-LR	r4i1p1
IPSL-CM5A-MR	r1i1p1
IPSL-CM5B-LR	r1i1p1
MIROC-ESM-CHEM	r1i1p1
MIROC-ESM	r1i1p1
MIROC5	r1i1p1
MIROC5	r2i1p1
MIROC5	r3i1p1
MPI-ESM-LR	r1i1p1
MPI-ESM-LR	r2i1p1
MPI-ESM-LR	r3i1p1
MPI-ESM-MR	r1i1p1
NorESM1-M	r1i1p1
inmcm4	r1i1p1

Table 1: Spliced historical-RCP8.5 simulations used

CNRM-CM5	r1i1p1
GFDL-CM3	r1i1p1
IPSL-CM5B-LR	r1i1p1
MIROC5	r1i1p1
NorESM1-M	r1i1p1
bcc-csm1-1	r1i1p1

Table 2: Model pre-industrial control simulation output used

CESM1-BGC	r1i1p1
CESM1-BGC	r1i1p2
CMCC-CM	r1i1p1
CNRM-CM5	r1i1p1
CSIRO-Mk3-6-0	r1i1p1
CanESM2	r1i1p1
FGOALS-s2	r1i1p1
GFDL-ESM2G	r1i1p2
GFDL-ESM2M	r1i1p1
GFDL-ESM2M	r1i1p2
IPSL-CM5A-LR	r1i1p1
IPSL-CM5A-MR	r1i1p1
MIROC5	r1i1p1
MPI-ESM-MR	r1i1p1
MPI-ESM-P	r1i1p1
MRI-CGCM3	r1i1p1
NorESM1-M	r1i1p1
bcc-csm1-1-m	r1i1p1
bcc-csm1-1	r1i1p1
inmcm4	r1i1p1

Table 3: CMIP5 1pctCO2 simulation output used

CCSM4	r1i1p10
CCSM4	r4i1p10
CCSM4	r6i1p10
CSIRO-Mk3-6-0	r1i1p4
CanESM2	r1i1p4
CanESM2	r2i1p4
CanESM2	r3i1p4
CanESM2	r4i1p4
GFDL-CM3	r1i1p1
GFDL-CM3	r5i1p1

Table 4: CMIP5 historicalMisc aerosol-only simulation output used