Drivers of Low-Frequency Sahel Precipitation Variability: Comparing CMIP5 and CMIP6 with Observations

Rebecca Herman[1], Michela Biasutti[2], and Yochanan Kushnir[3]

[1]Department of Earth and Environmental Sciences of Columbia University
[2]Lamont-Doherty Earth Observatory, Columbia University
[3]Lamont Doherty Earth Observatory of Columbia University

November 22, 2022

Abstract

We examine and contrast the simulation of Sahel rainfall in phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6). On average, both ensembles grossly underestimate the magnitude of low-frequency variability in Sahel rainfall. But while CMIP5 partially matches the timing and pattern of observed multi-decadal rainfall swings in its historical simulations, CMIP6 does not. To classify model deficiency, we use the previously-established link between changes in Sahelian precipitation and the North Atlantic Relative Index (NARI) for sea surface temperature (SST) to partition all influences on Sahelian precipitation into five components: (1) teleconnections to SST variations; the effects of (2) atmospheric and (3) SST variability internal to the climate system; (4) the SST response to external radiative forcing; and (5) the "fast" response to forcing, which is not mediated by SST. CMIP6 atmosphere-only simulations indicate that the fast response to forcing plays only a small role relative to the predominant effect of observed SST variability on low-frequency Sahel precipitation variability, and that the strength of the NARI teleconnection is consistent with observations. Applying the lessons of atmosphere-only models to coupled settings, we imply that the failure of coupled models in simulating 20th century Sahel rainfall derives from their failure to simulate the observed combination of forced and internal variability in SST. Yet differences between CMIP5 and CMIP6 Sahel precipitation do not mainly derive from differences in NARI, but from either their fast response to forcing or the role of other SST patterns.

Drivers of Low-Frequency Sahel Precipitation Variability: Comparing CMIP5 and CMIP6 with Observations

Rebecca Jean Herman,[a] Michela Biasutti,[b] Yochanan Kushnir[b]

[a] Department of Earth and Environmental Sciences of Columbia University, New York, NY

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

Corresponding author : Rebecca Herman, rebecca.herman@columbia.edu

ABSTRACT

We examine and contrast the simulation of Sahel rainfall in phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6). On average, both ensembles grossly underestimate the magnitude of low-frequency variability in Sahel rainfall. But while CMIP5 partially matches the timing and pattern of observed multi-decadal rainfall swings in its historical simulations, CMIP6 does not. To classify model deficiency, we use the previously-established link between changes in Sahelian precipitation and the North Atlantic Relative Index (NARI) for sea surface temperature (SST) to partition all influences on Sahelian precipitation into five components: (1) teleconnections to SST variations; the effects of (2) atmospheric and (3) SST variability internal to the climate system; (4) the SST response to external radiative forcing; and (5)

the "fast" response to forcing, which is not mediated by SST. CMIP6 atmosphere-only simulations indicate that the fast response to forcing plays only a small role relative to the predominant effect of observed SST variability on low-frequency Sahel precipitation variability, and that the strength of the NARI teleconnection is consistent with observations. Applying the lessons of atmosphere-only models to coupled settings, we imply that the failure of coupled models in simulating 20^{th} century Sahel rainfall derives from their failure to simulate the observed combination of forced and internal variability in SST. Yet differences between CMIP5 and CMIP6 Sahel precipitation do not mainly derive from differences in NARI, but from either their fast response to forcing or the role of other SST patterns.

1. Introduction

The semi-arid region bordering the North African Savanna and the Sahara Desert, known as the Sahel, received much scientific attention since it experienced unparalleled dramatic rainfall variability in the second half of the 20^{th} century. The importance of teleconnections between Sahel precipitation and global sea surface temperature (SST) was demonstrated in the early stages of Sahel climate variability research (Folland et al. 1986; Giannini et al. 2003; Knight et al. 2006; Palmer 1986; Zhang and Delworth 2006), and has been further reinforced in more recent studies (Okonkwo et al. 2015; Parhi et al. 2016; Park et al. 2016; Pomposi et al. 2015; Pomposi et al. 2016; Rodríguez-Fonseca et al. 2015 and references therein). But while the dominant role of SST in driving the pacing (though not necessarily the full magnitude) of 20th century Sahel rainfall variability is unquestioned (Biasutti 2019), there is still debate on whether the evolution of SST and the related Sahel precipitation variability were externally forced (Ackerley et al. 2011; Biasutti 2013; Biasutti and Giannini 2006; Biasutti et al. 2008; Bonfils et al. 2020; Dong and Sutton 2015; Giannini and Kaplan 2019; Haarsma et al. 2005; Haywood et al. 2013; Held et al. 2005; Hirasawa et al. 2020; Hua et al. 2019; Iles and Hegerl 2014; Kawase et al. 2010; Marvel et al. 2020; Polson et al. 2014; Undorf et al. 2018; Westervelt et al. 2017) or the manifestation of variability internal to the climate system (IV, Sutton and Hodson 2005; Ting et al. 2009; Zhang and Delworth 2006).

Recently, Herman et al. (2020, hereafter H20) investigated multi-model means (MMM) of historical simulations from the Coupled Model Intercomparison Project phase 5 (CMIP5, Taylor et al. 2012), and found that anthropogenic aerosols (AA) and volcanic aerosols (VA), but not greenhouse gases (GHG), were responsible for forcing simulated Sahelian precipitation that correlates well with observations, with AA alone responsible for the low-frequency component of simulated variability. This conclusion appeared consistent with previous claims that AA emissions, which increased until the 1970s and then decreased in response to clean air initiatives (Klimont et al. 2013; Smith et al. 2011), caused multi-decadal variability in Sahel precipitation via changes in Northern Hemisphere surface temperature (Ackerley et al. 2011; Haywood et al. 2013; Hwang et al. 2013; Undorf et al. 2018), or specifically via multidecadal variability in North Atlantic SST (the Atlantic Multidecadal Variability, AMV; Booth et al. 2012; Hua et al. 2019). However, H20 also found that the simulated rainfall response to forcing has little low-frequency power relative to observations, and that simulated IV is unable to account for this difference.

H20 and most other attribution studies do not examine in depth the pathways through which AA (and for that matter, IV and other external forcing agents) affect Sahel precipitation. Thus, H20 did not determine whether the discrepancy between CMIP5 simulations and observations represents an underestimate of aerosol indirect effects and climate feedbacks that amplify the simulated precipitation response to AA, or a fundamental inability of the models to simulate aspects of the observed climate response to forcing or observed modes of IV. Identifying the deficiencies in model representation of the pathways by which external forcing and IV influence the West African Monsoon and Sahel rainfall is essential for attribution of 20^{th} century changes and also for prediction of this region's climate future, as model simulations don't even agree on the sign of future precipitation changes in the Sahel (Biasutti 2013).

Here, we use the well-established link between SST and Sahel precipitation to decompose the effects of individual external forcing agents (F) and internal variability (IV) on Sahel precipitation (P) into five path components, presented in Figure 1: (1) teleconnections that communicate variations in SST to variations in P (indicated by the arrow $\overrightarrow{t}$); (2) the "fast" atmospheric and land-mediated effect of external forcing (F) on

P ($\overrightarrow{f}$); (3) the direct effect of atmospheric IV on P ($\overrightarrow{a}$); (4) the effect of F on SST ($\overrightarrow{s}$); and (5) the impact of IV in the coupled climate system on SST ($\overrightarrow{o}$). The path$F \to SST \to P$ is the "slow," SST-mediated effect of F on P.

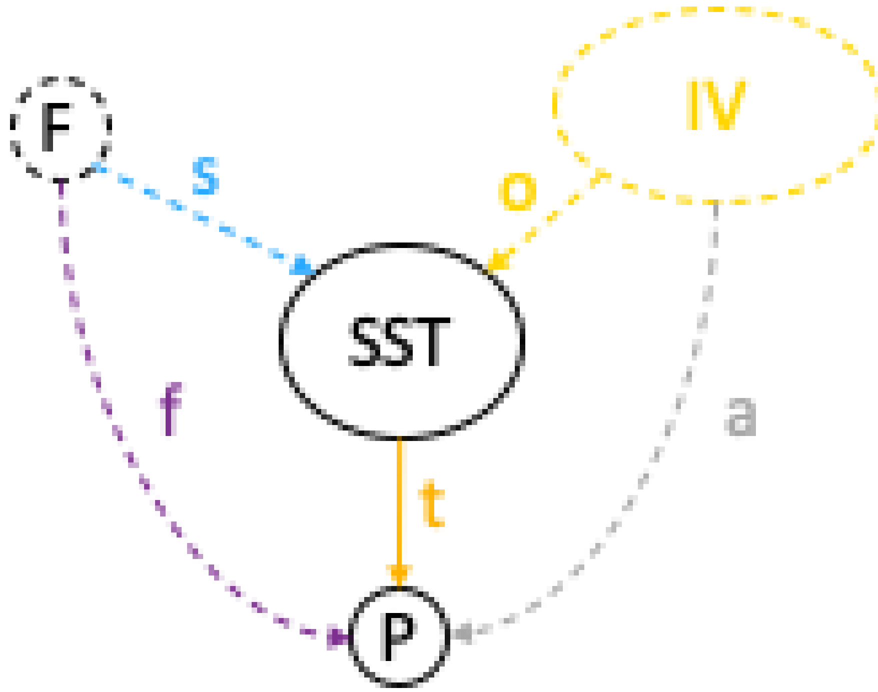

Fig. 1. Causal diagram relating external forcings (F), internal variability (IV), sea surface temperatures (SST), and Sahelian precipitation (P) via directional causal arrows. Unobserved variables and their causal effects are presented with dashed lines, while observed variables are presented with solid lines.

Characterization of these path components has been controversial. Firstly, separating the SST response to forcing ($\overrightarrow{s}$) from SST variability internal to the climate system ($\overrightarrow{o}$) has proven difficult (top of diagram). In particular, there is significant debate over whether observed AMV is a response to external forcing (Booth et al. 2012; Chang et al. 2011; Hua et al. 2019; Menary et al. 2020; Rotstayn and Lohmann 2002) or mainly an expression of IV in the Atlantic Meridional Overturning Circulation (AMOC, Han et al. 2016; Knight et al. 2005; Qin et al. 2020; Rahmstorf et al. 2015; Sutton and Hodson 2005; Ting et al. 2009; Yan et al. 2019; Zhang 2017; Zhang et al. 2016; Zhang et al. 2013) that is underestimated in models (Yan et al. 2018). This debate has been hard to resolve partially because IV in AMOC and aerosol forcing may have coincided by chance in the 20$^{\text{th}}$century (Qin et al. 2020). Next, examine the bottom of the diagram. The effect of the observed SST field on Sahel precipitation ($\overrightarrow{t}$) can be directly estimated using atmosphere-only simulations, but while these simulations capture the pattern of observed Sahel precipitation variability, many fail to capture its full magnitude (Biasutti 2019; e.g. Hoerling et al. 2006; Scaife et al. 2009). This could reflect an underestimate in climate models of the strength of SST teleconnections, which could be resolution dependent (Vellinga et al. 2016), or of land-climate feedbacks that amplify the teleconnections ($\overrightarrow{t}$), such as vegetation changes (Kucharski et al. 2013). But it could also reflect a significant additional role in the observations for a fast response to forcing ($\overrightarrow{f}$) that confounds the SST-forced signal[$P \leftarrow F \to SST \to P$; see Pearl et al. (2016) for notation] or coincides with it by chance.

To examine the path components in coupled simulations, we need a parsimonious characterization of the relationship between SST and Sahel precipitation. Giannini et al. (2013) and Giannini and Kaplan (2019, hereafter GK19) identify the North Atlantic Relative Index (NARI), defined as the difference between average

SST in the North Atlantic (NA) and in the Global Tropics (GT), as the dominant SST indicator of 20$^{\text{th}}$ century Sahel rainfall in observations and CMIP5 simulations. There are two main theories relating NARI to Sahelian precipitation (see Biasutti 2019; Hill 2019 for reviews of competing theories of monsoon rainfall changes). In the first, the "local view" (Giannini 2010), warming of GT causes even stronger warming throughout the tropical upper troposphere (Knutson and Manabe 1995; Parhi et al. 2016; Sobel et al. 2002), increasing thermodynamic stability across the tropics and inhibiting convection in an "upped ante" (Giannini and Kaplan 2019; Neelin et al. 2003) or "tropospheric stabilization" (Giannini et al. 2008; Lu 2009) mechanism. Warming of NA, on the other hand, is expected to thermodynamically increase moisture supply to the Sahel by increasing specific humidity over the NA, and thus destabilize the atmospheric column from the bottom up (GK19). The second theory interprets the relationship of Sahel precipitation to NARI, or, similarly, to the Atlantic meridional temperature gradient or the Interhemispheric Temperature Difference (ITD), as the result of an energetically-driven shift in the Intertropical Convergence Zone (ITCZ, Donohoe et al. 2013; Kang et al. 2009; Kang et al. 2008; Knight et al. 2006; Schneider et al. 2014) and the African rainbelt (e.g. Adam et al. 2016; Biasutti et al. 2018; Camberlin et al. 2001; Caminade and Terray 2010; Hoerling et al. 2006; Hua et al. 2019; Pomposi et al. 2015; Westervelt et al. 2017). According to both theories, an increase in NARI should wet the Sahel while a decrease causes drying. Given the prominence of the NARI teleconnection in the 20th century and the assumption of linearity, we approximate the full slow response as the product of the NARI response to external forcing and the strength of the NARI-Sahel teleconnection.

This paper is organized as follows: Section 2 provides details on the simulations and observational data used in this analysis while Section 3 discusses the methods. In Section 4.a, we update H20's analysis to the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al. 2016), examining the total response to forcing (all paths from F to P) and internal variability (all paths from IV to P). We then evaluate the performance of the CMIP6 AMIP simulations, decomposing them into the path components from the bottom half of Figure 1 ($\overrightarrow{t}$, $\overrightarrow{f}$, and $\overrightarrow{a}$) in Section 4.b, and focusing on the NARI teleconnection in Section 4.c. Section 4.d decomposes coupled simulations of NARI into the path components from the top half of Figure 1 ($\overrightarrow{s}$ and $\overrightarrow{o}$), while Section 4.e evaluates the consistency of the NARI teleconnection established in Section 4.c with coupled simulations. Finally, in Section 4.f, we use simulated NARI and the simulated NARI teleconnection to decompose the total response of Sahel precipitation to external forcing in coupled simulations (examined in Section 4.a) into fast and slow components. We discuss how our results fit in with the existing literature in Section 5 before concluding in Section 6.

2. Data

We examine coupled "historical" simulations from CMIP5 (Taylor et al. 2012) and CMIP6 (Eyring et al. 2016) forced with four sets of forcing agents—AA alone, natural forcing alone (NAT, which includes VA as well as solar and orbital forcings), GHG alone, and all three simultaneously (ALL)—as well as pre-Industrial control (piC) simulations, in which all external forcing agents are held constant at pre-Industrial levels. We additionally examine CMIP6 amip-piForcing (amip-piF) simulations, in which atmospheric models are forced solely with observed SST, and CMIP6 amip-hist simulations, which are forced with observed SST and historical ALL radiative forcing. Calculations with CMIP5 utilize the period between 1901 and 2003 while calculations with CMIP6 extend to 2014.

In H20, we used all available institutions for each forcing subset. Here, in order to provide a more stringent comparison of the effects of different forcing agents, we exclude institutions from the coupled ensemble that do not provide AA, GHG, and ALL simulations, and from the AMIP ensemble if they do not provide both amip-piForcing and amip-hist simulations. We additionally exclude piC simulations that are shorter than the historical simulations as well as any simulations with data quality issues. Tables S1-S3 enumerate the simulations used in this analysis.

Precipitation observations are from the Global Precipitation Climatology Center (GPCC, Becker et al. 2013) version2018, and SST observations are from the National Oceanic and Atmospheric Administration's (NOAA) Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5, Huang et al. 2017).

We analyze precipitation over the Sahel (12°-18°N and 20°W-40°E) and the SST indices of GK19: the North Atlantic (NA, 10°-40°N and 75°-15°W), the Global Tropics (GT, ocean surface in the latitude band 20°S-20°N), and the North Atlantic Relative index (NARI, the difference between NA and GT). All indices are spatially- and seasonally-averaged for July-September (JAS).

3. Methods

The multi-model mean (MMM) for a set of simulations consists of a 3-tiered weighted average over (1) individual simulations (runs) from each model, (2) models from each research institution, and (3) institutions in that ensemble. Details of the weighting are provided in H20; the results are robust to differences in weighting. Time series are not detrended, and anomalies are calculated relative to the period 1901-1950.

To evaluate the performance of the simulations relative to observations, we compute correlations (r), which capture similarity in frequency and phase, and root mean squared errors standardized by observed variance (sRMSE), which measure yearly differences in magnitude between the simulations and observations. An sRMSE of 0 represents a perfect match between simulations and observations, and 1 would result from comparing the observations with a constant time series.

To estimate uncertainty in the forced MMMs and associated metrics, we apply a bootstrapping technique to the last tier of the MMM as described in H20, yielding a probability distribution function (pdf) about the MMM and each metric. Due to the finite number of simulations, these pdfs underestimate the true magnitude of the uncertainty. We evaluate significance by applying a randomized bootstrapping technique, which increases the effective sample size, to the piC simulations with one significant improvement over H20: instead of using just one subset of each piC simulation at a random offset in the first tier of the MMM in each bootstrapping iteration, we take enough subsets to match the number of that model's historical runs. Done this way, the confidence intervals calculated using piC simulations accurately represent noise in the forced MMMs. PiC pdfs from the same ensemble differ slightly because many institutions provide a different number of simulations for different subsets of forcing agents (see Table S2). Where the piC pdfs and confidence intervals are similar enough, they are presented together with a single grey dotted curve and dashed line; when they differ, they are presented in the colors associated with the relevant forcings.

We perform a residual consistency test, which compares the power spectra (PS) of individual simulations to that of observations, with one significant modification over H20: we calculate the PS using the multi-taper method. Confidence intervals for the PS for observations and MMMs are given by the multi-taper method, without accounting for the uncertainty in the MMMs themselves. Mean PS by model are colored by climatological rainfall bias given by those simulations. The multi-model mean of these PS, or the "tiered mean", is calculated using the three tiers from the definition of the MMM, but without weights, since spectral power is not attenuated when averaging PS.

4. Results

a. Changes in CMIP6: Total Precipitation Response to Forcing and Internal Variability

If Sahelian precipitation is a linear combination of IV in the coupled climate system and variability forced by external agents, then the MMM over coupled simulations with differing initial conditions filters out atmospheric and oceanic IV ($\overrightarrow{a}$ and $\overrightarrow{o}$), leaving the fast and slow precipitation responses to external radiative forcing ($\overrightarrow{f}$ and $F \rightarrow SST \rightarrow P$). Figure 2 compares observed Sahelian precipitation anomalies (black, left ordinates) to the MMM anomalies of simulated Sahelian precipitation (right, amplified colored ordinates) in CMIP5 (dotted curves) and CMIP6 (solid curves) for four sets of forcing agents: ALL (a, blue), AA (b, magenta), natural forcing (c, "NAT," brown and red), and GHG (d, green). The figure also presents the bootstrapping 95% confidence intervals of the forced CMIP6 MMMs (blue, magenta, brown, and green shaded areas) and of MMMs over the CMIP6 piC simulations (yellow shaded areas) on the right ordinates. The width of the yellow shaded areas represents the magnitude of noise deriving from coincident IV in the MMMs. Differences in its width between panels arise from varying numbers of simulations for the different forcing subsets (see Methods and Table S2).

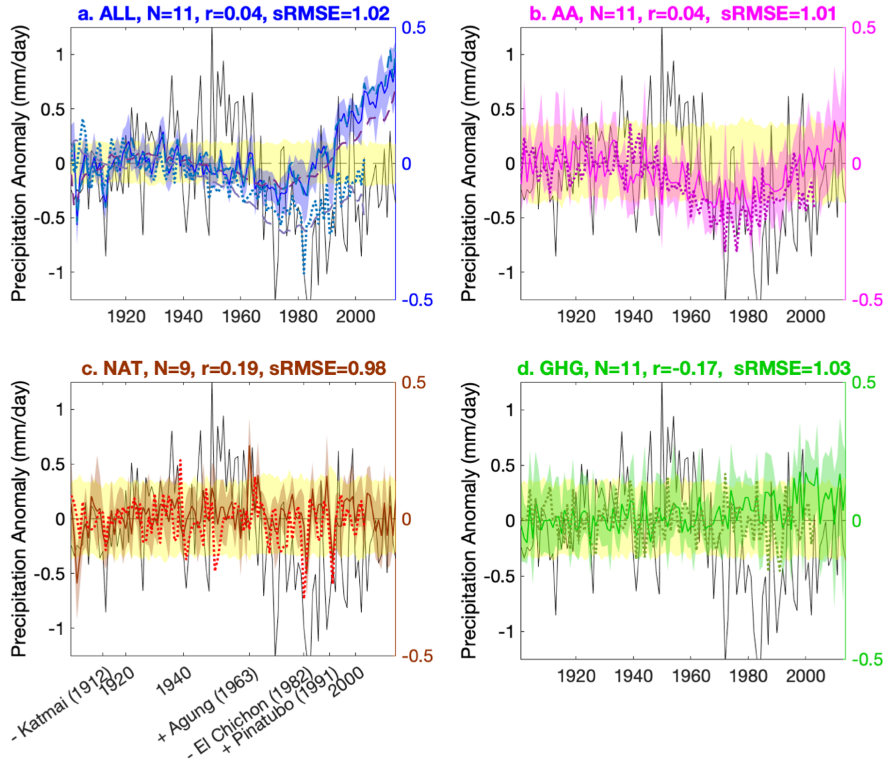

Fig. 2. Observed (black, left ordinates) and simulated (colored, right ordinates) Sahelian precipitation anomalies, forced with ALL (a, blue), AA (b, magenta), NAT (c, brown/red), and GHG (d, green). The CMIP6 MMMs are presented with solid curves surrounded by shaded areas demarking the bootstrapping confidence interval, while the CMIP5 MMMs are presented with dotted curves. The yellow shaded area is the confidence interval of randomized bootstrapped MMMs of CMIP6 piC simulations, and represents the magnitude of noise in the CMIP6 MMMs. Hemispherically asymmetric volcanic forcing from Haywood et al noted in panel (c). A negative sign denotes an eruption that cooled the northern hemisphere more than the southern hemisphere while a positive sign denotes the opposite, aligning with the sign of the expected Sahelian precipitation response to the eruption. Panel (a) additionally shows the CMIP6 ALL MMM when restricted to models, rather than institutions, that provide AA simulations (blue dashed curve), and a 20-year running mean of the sum of the AA, NAT, and GHG MMMs for CMIP5 (lavender dashed curve) and CMIP6 (burgundy dashed curve). The label shows the number of institutions used for each CMIP6 MMM (N), the correlation of the CMIP6 MMM with observations (r), and the standardized root mean squared error of the CMIP6 MMM with observations (sRMSE).

In the AA experiments (panel b), CMIP6 is anomalously wetter than CMIP5 in the 1970s and around 2000, but otherwise looks similar to CMIP5: precipitation declines in the mid-century and then recovers after the clean air acts, preceding the timing of observed variability by about 10 years. There are some differences in the NAT experiments between CMIP5 and CMIP6 (panel c), but the largest variations in both ensembles are interannual episodes that are clearly associated with volcanic eruptions. In the GHG experiments (panel d), CMIP6 shows anomalous wetting after 1970 that wasn't present in CMIP5.

Similar changes can be seen in the ALL simulations (panel a): while CMIP5 reaches peak drought in 1982 – close to the observed precipitation minimum – CMIP6 dries very little and only until 1970, after which it displays an anomalously wetter climate than CMIP5 through the end of the century. But while the precipitation responses to different forcing agents appear to add linearly in CMIP5 (compare the lavender dashed curve to the blue dotted curve), the late century wetting in CMIP6 is larger than the sum of GHG and AA wetting (burgundy dashed curve; including NAT does not help.) This effect is robust to differences in model availability for the different sets of forcing agents (see figure caption and light blue dashed curve). Thus, in the ALL simulations, CMIP6 displays slightly less drying from AA compared to CMIP5, more wetting from GHG, and additional wetting after 1990 from a non-linear interaction between forcings.

As a result of these changes, the response to forcing in CMIP6 is a poor match to observations. Figure 3 displays the correlation (panel a, "r") and sRMSE (panel b) between observations and simulated MMMs (dots) and bootstrapped MMMs (curves) from CMIP6 (ALL in blue, AA in magenta, NAT in brown, and GHG in green solid curves) and CMIP5 (ALL and AA in blue and magenta dotted-dashed curves; other simulations omitted for clarity) from 1901 to the end of the simulations (2003 for CMIP5 and 2014 for CMIP6). The dotted curves present the randomized bootstrapping distributions for the CMIP6 piC simulations, and the vertical dashed lines mark the one-sided p=0.05 significance level given by these distributions. Recall that correlation measures similarity in timing between simulations and observations where 1 is a perfect match, and sRMSE measures the amplitude of differences between the simulations and observations where 0 is a perfect match.

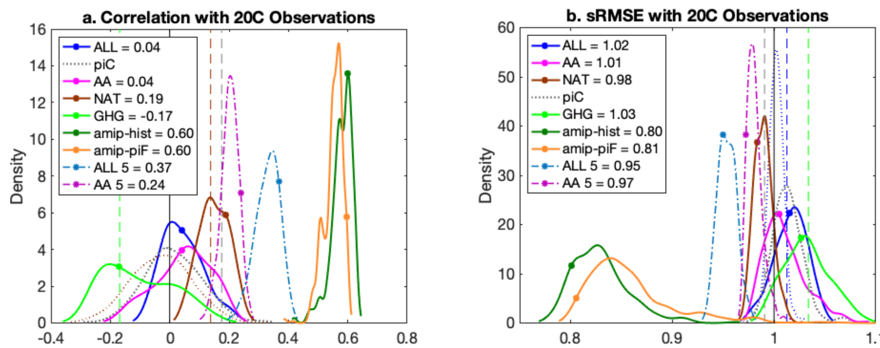

Fig. 3. Correlations (a) and standardized RMSE (b) between observations and historical and AMIP simulations from CMIP6 (1901-2014, solid) and those simulations from CMIP5 that outperform the CMIP6 historical simulations (1901-2003, dotted-dashed, legend entries include "5"). Dots and stars denote the statistic between the MMM and observations, while the curves denote the bootstrapping pdfs. The dotted grey curves display the bootstrapping pdfs for the same statistics applied to a MMM over the CMIP6 piC simulations, and the grey dashed lines mark the one-sided p=0.05 significance level given by the piC distribution. Colored dotted curves and dashed lines show the piC distributions associated with those subsets of forcing agents for which the piC distribution differs noticeably from those of the other subsets of forcing agents.

CMIP5's AA (r = 0.24, sRMSE = 0.97) and ALL (r = 0.37, sRMSE = 0.95) MMMs achieve significance in both metrics – a fact that, in isolation, is consistent with the suggestion that AA may explain observed variability but underestimate its magnitude. Instead, in CMIP6, AA (r = 0.04, sRMSE = 1.01) and ALL (r = 0.04, sRMSE = 1.02) do not perform statistically better than noise, and GHG performs significantly worse (r = -0.17, sRMSE = 1.03). The additional years included in the CMIP6 simulations (2004-2014) cannot explain the entire deterioration of performance between CMIP5 and CMIP6: even when restricted to CMIP5's time period, CMIP6 ALL and AA simulations both perform worse than CMIP5 in both metrics (r = 0.07 and sRMSE = 1.00 for AA, r = 0.13 and sRMSE = 0.99 for ALL). Most of the remaining deterioration in performance for AA is due to reduced drying in the 1970s in CMIP6. In CMIP6, NAT (r = 0.19, sRMSE = 0.98) is the only forcing that performs significantly well. We conclude that aside from episodic responses to volcanic eruptions, the ensemble of coupled CMIP6 simulations has no significant skill in simulating historical Sahel rainfall in response to external forcing.

As in CMIP5, the simulated forced component of precipitation changes in CMIP6—given by the MMM—has a much smaller variance than observations (note the amplification of the right ordinates in Figure 2).

However, the poor performance of the CMIP6 simulations makes it clear that amplifying the simulated forced component will not help explain observed precipitation.

For simulated atmospheric and oceanic IV ($\overrightarrow{a}$ and $\overrightarrow{o}$) to explain observed precipitation variability, it is not enough that observed yearly Sahelian precipitation anomalies fall within the range of individual simulations (not shown)—the latter must also match the distinctive low-frequency power of the observations. In Figure 4 we compare the power spectra (PS) of piC simulations (colored brown to turquoise by model climatological rainfall) to the observed PS (solid black) and the PS of the ALL-residual (observations minus the ALL MMM, dotted-dashed black). In the observations and the residual, variance at periods longer than about 20 years (low-frequency) is roughly 5 times as large as the high-frequency variance. Low-frequency variability in the piC simulations is smaller than, and inconsistent with, either observed or residual variability. Moreover, it is similar in magnitude to simulated high frequency variability, suggesting that IV in simulated Sahel rainfall derives mostly from atmospheric ($\overrightarrow{a}$), rather than oceanic ($\overrightarrow{o}$), IV, or that simulated oceanic IV is too white (Eade et al. 2021). Because the shape of the spectrum is wrong, even a bias correction that inflates simulated internal variability would not bring simulations and observations into alignment.

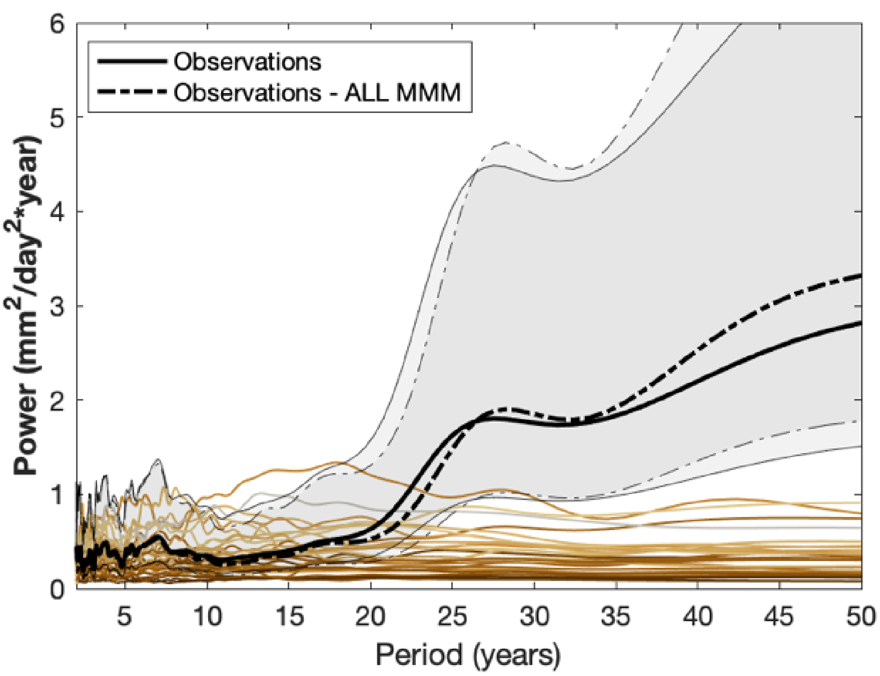

Fig. 4. PS of observed Sahelian precipitation (solid black curve) and the residual of observations and the ALL MMM (dotted-dashed black curve) and associated 95% confidence intervals (grey shading), compared to the average PS by model of piC simulations (brown to turquoise). Mean piC PS are colored by the average yearly piC precipitation by model, where brown simulations are drier than observed, and turquoise simulations are wetter than observed.

We must conclude that no linear combination of the simulated forced signal (which correlates poorly with observations) and simulated IV (which has insufficient low-frequency variance) in coupled CMIP6 simulations can explain observed Sahel variability during the 20$^{\text{th}}$ century. Thus, model deficiency cannot be blamed solely on the simulation of climate feedbacks: the CMIP6 ensemble displays a fundamental inability to simulate the observed fast and slow Sahelian precipitation responses to forcing, observed low-frequency IV, or both. To identify the proximate cause of this failure, in the next three sections we examine each causal path component identified in Figure 1.

b. AMIP simulations: the Response to SST, Atmospheric Internal Variability, and the Fast Response to

Forcing ($\overrightarrow{t}$, $\overrightarrow{a}$, and $\overrightarrow{f}$)

To isolate the effect of SST on the Sahel ($\overrightarrow{t}$), we examine precipitation in the CMIP6 amip-piForcing simulations, which force atmosphere-only models with the observed SST history (containing both internal, $\overrightarrow{o}$, and forced, $\overrightarrow{s}$, oceanic variability) and constant preindustrial external radiative forcing (no $\overrightarrow{f}$). The MMM of simulated Sahel precipitation filters out atmospheric IV ($\overrightarrow{a}$), leaving the precipitation response to the entire observed SST field. It is displayed in Figure 5a (orange) and compared to observations (black) on the same ordinates. Overall, the performance of the amip-piF MMM is much better than that of the coupled simulations: it achieves a high correlation (r = 0.60) and a low sRMSE (0.81, see orange curves in Figure 3). The good match with observations is achieved mostly at low frequencies: though it doesn't accurately capture many interannual episodes—notably including the precipitation minimum in 1984—the MMM appears to capture the magnitude of low-frequency variability, even including wetting in the 50s and early 60s, which is missing from the coupled MMM. This can be seen more quantitatively by spectral analysis. In Figure 6a, the PS of the amip-piF MMM (dashed orange curve) and its 95% confidence interval (orange shaded areas), are compared to those of observations (black). Unlike previous generations of AMIP experiments (e.g. Scaife et al. 2009), the PS of the simulated MMM is roughly consistent with observations.

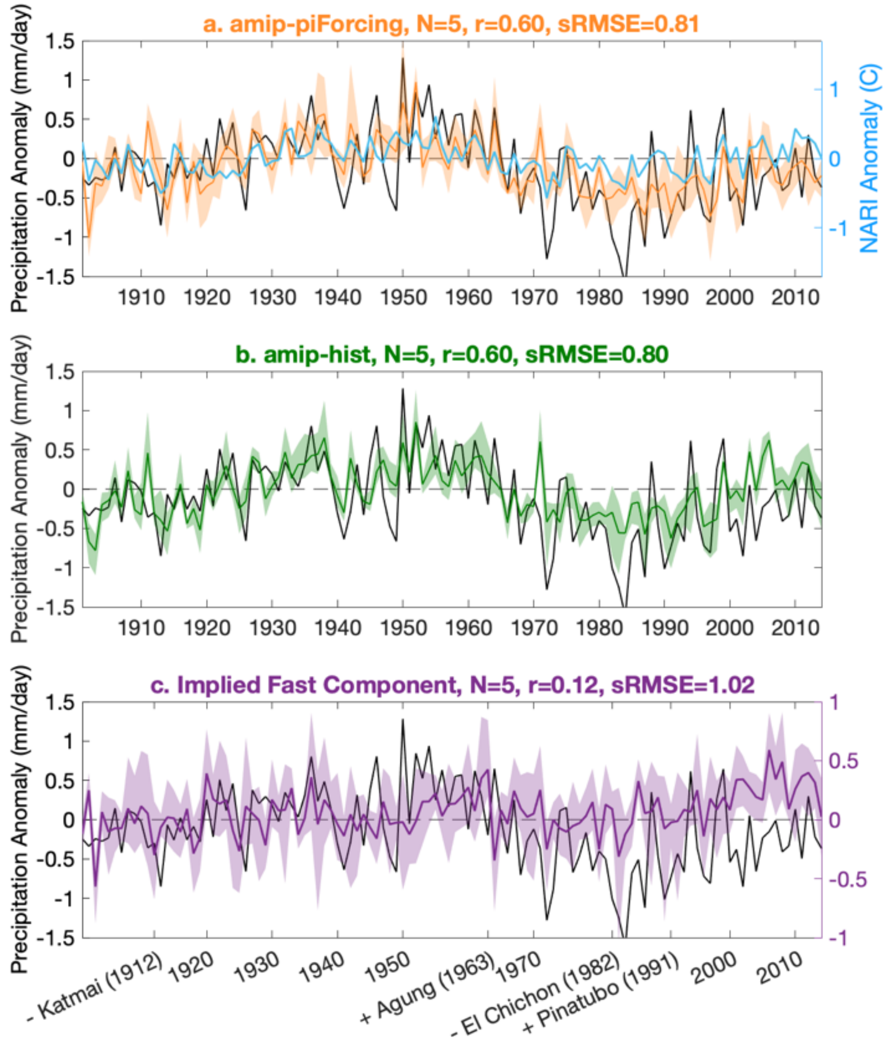

Fig. 5. Observed (black) and simulated (colored) Sahelian precipitation anomalies, forced with observed

SST alone (a, amip-piF, orange) and with observed SST and all external forcing agents (b, amip-hist, dark green). The shaded areas denote the bootstrapping confidence intervals about the simulated MMMs. Panel (a) additionally displays observed NARI (light blue, right ordinates). The right ordinates for panel (a) are scaled by the inverse of the simulated amip-piF teleconnection strength (see Section 4.c) so that when read on the left ordinates, NARI represents its predicted impact on precipitation. Panel (c) compares observed precipitation (left ordinates) to the implied simulated fast component in AMIP simulations (amip-hist − amip-piF, purple, right ordinates). As in Figure 2, panel (c) denotes hemispherically asymmetric volcanic eruptions, where the sign denotes the sign of the expected Sahelian precipitation response to the eruption.

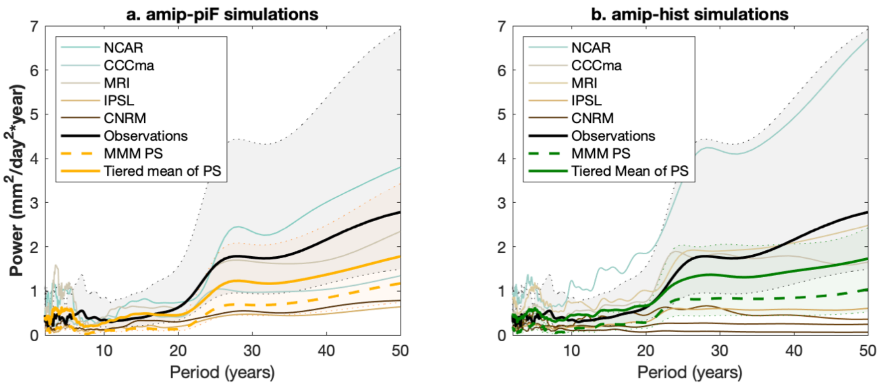

Fig. 6. PS of observed Sahelian precipitation (black) and associated 95% confidence interval (black shading) compared to the PS of amip-piF simulations (a) and amip-hist simulations (b). As in Figure 4, mean PS by model are colored by average yearly precipitation, where brown is drier than observed, grey is observed, and turquoise is wetter than observed. The mean PS across models is displayed in orange for amip-piF (a) and in green for amip-hist (b). The dashed lines show the PS of the MMMs with associated 95% confidence intervals (colored shaded areas).

The curves colored brown to turquoise in Figure 6 show the average by model of the PS of individual simulations, colored by climatological Sahelian precipitation bias. We note that wet-biased simulations (turquoise) have more power than dry-biased simulations (brown), consistent with the expected relation between the mean and variance of precipitation. The tiered mean over these PS is presented in solid orange; it contains atmospheric IV ($\overrightarrow{a}$) in addition to SST-forced variability ($\overrightarrow{t}$). Though it is not statistically different from the MMM PS, atmospheric white noise gives it slightly more power at all frequencies, and thus it is clearly consistent with the observed PS (black). Global SST forcing, while unable to explain much of observed high frequency variability in Sahelian precipitation (note the low power of the dashed orange curve at periods below 20 years), is able to reproduce the pattern and, in combination with atmospheric IV, the full magnitude of observed multi-decadal precipitation variability.

We now estimate the “fast” precipitation response to ALL in the CMIP6 AMIP simulations (Figure 5c, purple, $\overrightarrow{f}$) by subtracting the MMM of amip-piF simulations (a, orange) from that of amip-hist simulations (b, green), the latter of which are forced with historical SST and historical external radiative forcing. The AMIP “fast” MMM shows some episodic variability that is consistent with the coupled NAT MMM, and a wetting trend after 1985. On its own, it is only weakly correlated to observations (r = 0.12, sRMSE = 1.02), and it has relatively low amplitude. When combined with SST forcing in the amip-hist simulations, it has little effect: correlation stays at 0.60 and sRMSE is reduced from 0.81 only to 0.80 (compare green and orange curves in Figure 3) and spectral properties are virtually unchanged (Figure 6). The best linear fit to observed precipitation would combine the amip-piF MMM with the fast response to forcing scaled down by a factor of 0.3 ± 0.2. The fast response may be overestimated in AMIP simulations because the radiative forcing has directly contributed to generating observed SST which is prescribed in the simulations, and because the magnitude of the radiative forcing itself may be overestimated, as suggested by Menary et al. (2020).

The high performance of the amip-piF simulations and the small impact of the potentially overestimated fast response to forcing suggest that the principal deficiency in simulating low-frequency Sahelian precipitation variability in coupled models stems from a deficiency in simulating the observed combination of forced and internal variability in SST, and not from a failure to reproduce the observed teleconnection strength or fast response to forcing.

c. The NARI Teleconnection: AMIP Simulations and Observations ($\overrightarrow{t}$)

We next determine the strength of the linear NARI-Sahel teleconnection and investigate how well it represents the effect of global SST on Sahel precipitation in simulations and observations. Observed NARI anomalies relative to the 1901-1950 mean are presented in Figure 5a in light blue on the right ordinates. NARI correlates well with SST-forced Sahelian precipitation in the amip-piF simulations (orange, left ordinates;$r = 0.52 \pm 0.10,\ r = 0.60\ for\ the\ actual\ MMM$), but still leaves 64% of its variance unexplained, suggesting influences from other SST patterns or non-linear or non-stationary effects (Losada et al. 2012). Some of the unexplained variance is at faster timescales than those of our interest, but not all. Let's assume that the influences of NARI and other ocean basins on Sahel precipitation are linear and add linearly, and that the NARI teleconnection is unconfounded by the influence of other ocean basins; then we can measure the strength of the NARI teleconnection by the regression coefficient of the amip-piF precipitation MMM, which contains only SST-forced variability, on NARI. This calculation yields a regression slope of $0.87 \pm 0.26 \frac{\text{mm}}{day*C}$. This value is affected by both high- and low-frequency variability, which is appropriate if the teleconnection is, indeed, linear. The left ordinates in Figure 5a are scaled relative to the right ordinates by this teleconnection strength so that, when read on the left ordinates, the light blue curve represents the expected precipitation response to NARI. This view highlights how NARI captures the timing of simulated low-frequency variability, even though it fails to explain the full magnitude of simulated dry anomalies after 1975. In the rest of this paper we use the NARI teleconnection as the best linear representative of the simulated influence of SST on Sahel precipitation in the 20th century.

The teleconnection strength calculated from the amip-piF simulations is not directly comparable to observations, because the latter includes the fast precipitation response to forcing, which can confound estimates of the teleconnection. A comparison can be drawn between the apparent teleconnection strength in the amip-hist simulations (0.93 ± 0.41) and in observations (1.04). The consistency lends credence to our previous suggestion that simulated SST teleconnections to Sahel rainfall appear to have the appropriate strength in CMIP6, at least in the amip simulations.

d. Forced and Internal SST Variability in Coupled Simulations ($\overrightarrow{s}$ and $\overrightarrow{o}$)

We now examine simulation of forced ($\overrightarrow{s}$) and internal ($\overrightarrow{o}$) SST variability. Figure 7 compares observations (black) to the simulated SST response to forcing ($\overrightarrow{s}$)—represented by MMM anomalies (colors)—for NARI (right column) and its constituent ocean basins – the North Atlantic (NA, left column) and the Global Tropics (GT, middle column). The yellow shaded areas show the bootstrapping 95% confidence intervals of the piC simulations for statistical significance, while the other shaded areas denote uncertainty in the CMIP5 and CMIP6 MMMs. As above, CMIP5 MMM anomalies are presented in dotted curves and CMIP6 in solid curves, color-coded according to their forcing.

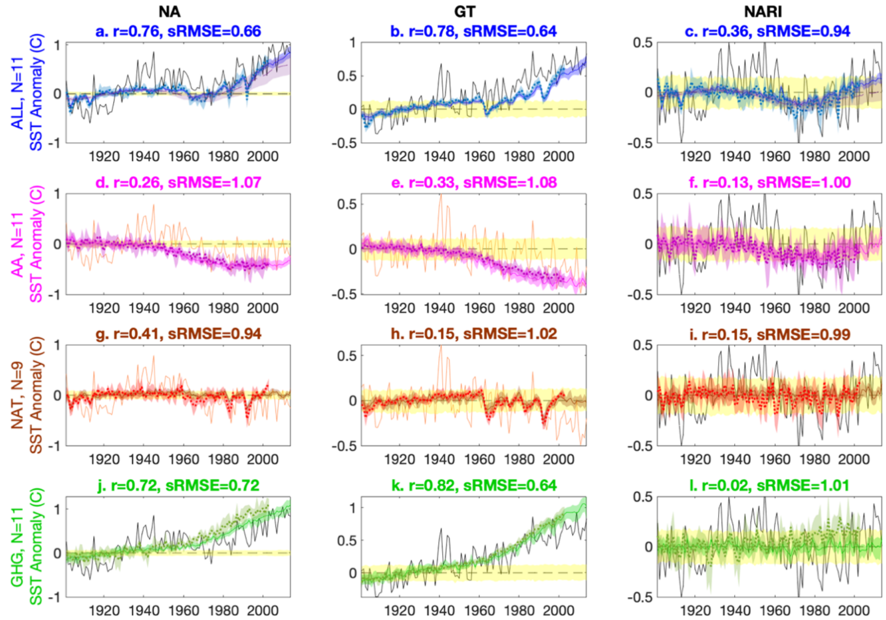

Fig. 7. Observed (black) and simulated CMIP5 and CMIP6 SST anomalies (relative to 1901-1950) for the North Atlantic (NA, left column), the Global Tropics (GT, middle column), and the North Atlantic Relative Index (NARI, right column) when forced with ALL (blue, top row), AA (magenta, second row), NAT (brown/red, third row), and GHG (green, bottom row). The CMIP6 MMMs are presented with solid curves while the CMIP5 MMMs are presented with dotted curves. Both are surrounded by shaded areas demarking the bootstrapping confidence interval. Panels (a) and (c) additionally display a 20-year running mean of the sum of simulated NA and NARI over the individual forcing simulations for CMIP6 (burgundy dashed curve) with associated bootstrapping confidence interval (burgundy shaded area). Including NA in the sum makes little difference. For NA and GT under AA and NAT (middle two rows and left two columns), the orange curve displays detrended observations, calculated by subtracting simulated GHG-forced SST (bottom row) from observations in that ocean basin. The yellow shaded area is the confidence interval when bootstrapping the MMM of CMIP6 piC simulations, and represents the magnitude of noise in the CMIP6 MMMs. A horizontal black dashed line marks 0 anomaly, which represents the average SST from 1901-1950. The y labels show the number of institutions that were used for each subset of forcing agents in CMIP6 (N, see Table S2), and the subplot titles display the correlation (r) and sRMSE between the MMM and observations for CMIP6.

Observed NARI (panel c, black) shows strong multi-decadal variability throughout the century. In the ALL simulations (top row, blue), the temporal evolution of NARI (c) matches the observations with some skill (r=0.40, sRMSE = 0.92 for CMIP6), but fails to capture the full magnitude of observed cooling in the 1970s and 80s or, more prominently, any multi-decadal variability prior to 1960. Moreover, its GT and NA components do not match very well either the observed, roughly linear warming trend in GT (b), or the marked multi-decadal variability in NA (a). In both CMIP5 and CMIP6 ALL simulations, the simulations of GT (b, blue) are anomalously colder than observations between 1960 and 2000, when simulated AA cooling (e, magenta) is the strongest and not yet compensated by GHG warming (k, green), leading us to question whether the match of simulated and observed NARI in this period happens due to compensating errors. For NA, the match between observations and the ALL-forced response is better in the later part of the record, but worse in the first half. During the period prior to 1960, according to both CMIP ensembles, GHG warming (j, green) masks AA cooling (d, magenta) to produce a roughly constant temperature in the ALL simulations (a, blue). The simulated cold episode in 1964 is due to the eruption of Agung in 1963 (g, brown and red), and it is only after the mid 1960's that increased GHG warming overtakes stagnating AA cooling to produce pronounced warming in fairly good accord with observations. Much of the observed variability in NA (a, black) thus does not seem to be a response to external radiative forcing.

The AA forcing had appeared to explain observed low-frequency Sahel precipitation variability in H20, but we now see that it might be the right result for the wrong reason. AA (second row, magenta) produce low-frequency NARI variability (f), but this simulated NARI is a poor match to observations (f, r=0.10, sRMSE = 1.04 for CMIP5; r=0.07, sRMSE=1.09 for CMIP6; a performance statistically worse than noise). The difference between simulations and observations is even more stark in NARI's constituent ocean basins. We can attempt to compare AA-forced NA and GT to an observed "GHG-residual" (that is, the observation minus the GHG-forced MMM, presented in orange instead of black), which represents our best estimate of the sum of observed oceanic IV and the observed responses to aerosols. This index shows marked, roughly stationary low-frequency variability in NA (d, orange), which contrasts with a more monotonic behavior in the simulated NA index (magenta). In particular, we note that the AA simulations display an especially steep decline in NA SST between ˜1940 and 1980, but monotonic cooling throughout the century. Though legislation to curb pollution reduced AA loading in the northern hemisphere after 1970 (Hirasawa et al. 2020), simulated NA doesn't warm at all before 2010. Overall, the effect of reducing AA emissions in both CMIP ensembles is to halt the cooling of NA, not to cause actual warming. This is consistent with estimates of the hemispheric difference in total absorbed solar radiation in AA simulations in CMIP6, which level off, but do not decrease, after 1970 (Menary et al. 2020).

Could internal SST variability ($\overrightarrow{o}$) explain the difference between the simulated response to forcing and observations in these ocean basins? In Figure 8, we present the mean PS of SST for piC simulations from each CMIP6 model (colder than observed models are in blue and warmer than observed models are in red). We compare these PS to the PS for observed SST (solid black), the GHG-residual (dotted-dashed black), and/or the ALL-residual (dotted black), avoiding time series with dramatic trends (see subplot legends). Simulated IV in most of the CMIP6 models used in this study does not match residual or observed low-frequency variability in NA (a), GT (b), or NARI (c). In CMIP5, SSTs are colder and IV at all frequencies is larger than in CMIP6, but no model shows an increase in spectral power at low frequencies for any SST index (not shown). There are, however, three CMIP6 models for which low-frequency IV in NA is not inconsistent with model physics: CNRM-ESM2-1 p1 (pink), IPSL-CM6A-LR p1 (blue), and CNRM-CM6-1 p1 (grey). Certainly, either the simulated SST response to forcing, simulated oceanic internal variability, or both, are not well represented in the CMIP ensembles, and this is the primary reason that coupled CMIP simulations cannot reproduce observed 20th century Sahel rainfall.

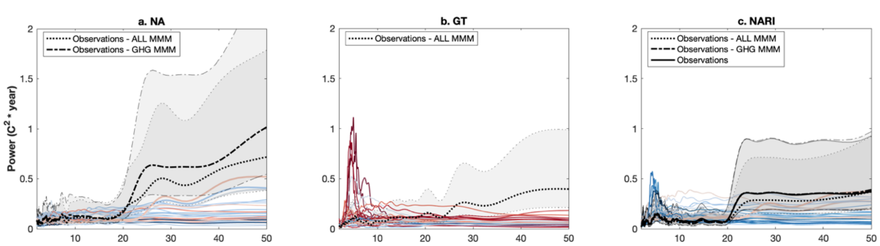

Fig. 8. PS of observed SST (solid black), observed SST – GHG MMM (dotted-dashed black), observed SST – ALL MMM (dotted black) and associated 95% confidence intervals (black shading) in NA (a), GT (b), and NARI (c), compared to the PS of piC simulations. Similar to Figure 4, mean PS by model are colored by average SST, where blue is colder than observed, grey is observed, and red is warmer than observed.

However, deficiencies in simulating SST cannot explain the difference in simulated externally forced precipitation variability between CMIP5 and CMIP6. The only notable difference in simulated SST between the two ensembles is that CMIP6 warms NA (and therefore NARI) less than CMIP5 in the GHG simulations (Figure 7j and l). As in simulated Sahel precipitation, warming of NA and NARI in CMIP6 ALL simulations is larger than the smoothed sum of simulated SST change in the individual-forcing simulations (burgundy dashed curve), which, aside from volcanic eruptions, remains below the confidence interval for the CMIP6 MMM (dark blue shaded area) from 1950 onward (this discrepancy is, again, robust to differences in model availability for the different sets of forcing agents). Thus, a non-linear interaction between forcing agents in CMIP6 balances the additional SST warming in CMIP5 in the ALL simulations, and the difference in coupled simulations of Sahel rainfall between CMIP5 and CMIP6 must derive from changes in the fast response to forcing, SST teleconnections, or both.

e. The NARI teleconnection in Coupled Simulations

Now that we have examined SST in the coupled simulations, we may determine whether the teleconnection strength estimated from amip-piF simulations is consistent with coupled simulations. This is verified by the fact that the amip-piF teleconnection strength falls within the range of teleconnection strengths calculated from individual piC simulations in CMIP5 (0.5 ± 0.6) and CMIP6 (0.4 ± 0.6), but the ranges are large (possibly because the increased presence of atmospheric and oceanic IV and decreased variance of NARI in the individual piC simulations obscures the teleconnection). As a second test, we compare the confounded teleconnection strength in the amip-hist simulations (0.93 ± 0.41) to that of bootstrapped MMMs in the coupled ALL simulations in CMIP5 (0.66 ± 0.28) and CMIP6 (1.5 ± 0.3). The confounded teleconnection strength in amip-hist simulations is consistent with the confounded estimate in CMIP5, but is smaller than and inconsistent with the confounded estimate in CMIP6. This may be because NARI variability in the coupled simulations is smaller relative to the magnitude of external radiative forcing than it is in the amip-hist simulations. If this is the cause for the apparent inconsistency, we may still confirm the NARI teleconnection strength in CMIP6 simulations by showing that the implied fast response to forcing is consistent with the fast response from the amip-hist simulations.

f. Fast and Slow Responses to Forcing in Coupled Simulations ($\overrightarrow{f}$ and$F \rightarrow SST \rightarrow P$)

Under the assumption that the dominant simulated path of SST influence on the Sahel is captured by a linear relationship with NARI, we estimate the slow response to forcing in coupled simulations as the simulated NARI MMM scaled by the teleconnection strength derived from uncoupled simulations ($0.87 \frac{\text{mm}}{day\ C}$, Section 4.c), so that a warm (cold) NARI predicts a wet (dry) Sahel. In Figure 9, simulated NARI (as in Figure 7, right column) is displayed on the left ordinates in light blue (CMIP6) and turquoise (CMIP5). The right ordinates are scaled by the teleconnection strength so that, when read on the right ordinates, simulated NARI represents the estimated slow component of the precipitation response to forcing. Also on the right ordinates are the total simulated precipitation responses to forcing (as in Figure 2) in CMIP5 (right column) and CMIP6 simulations (left column), colored by forcing agents. The simulated precipitation responses to forcing (colors) match the estimated slow response to forcing (turquoise) reasonably well: the main differences appear after about 1970 in CMIP5 and 1990 in CMIP6.

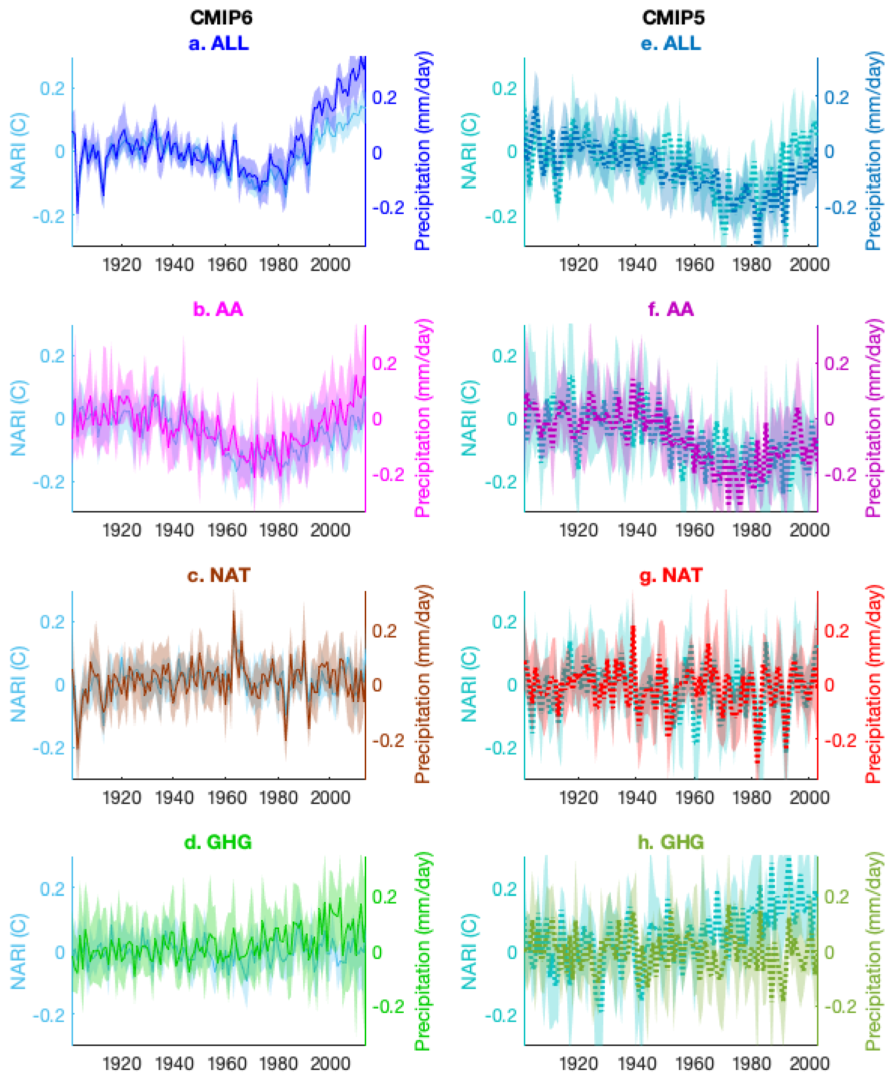

Fig. 9. Simulated Sahel precipitation (right ordinates, same as Figure 2) MMMs (solid and dotted curves) and associated 95% confidence intervals (shaded areas) in CMIP5 (right column) and CMIP6 (left column) when forced with ALL (blue, top row), AA (magenta, second row), NAT (brown/red, third row), and GHG (green, bottom row), compared to simulated NARI (left ordinates, light blue and turquoise, same as Figure 7). The right ordinates are scaled such that a 1°C change in NARI corresponds to a 0.87 mm/day change in precipitation, given by the teleconnection strength in the CMIP6 amip-piF simulations (see Section 4.c).

We expect the differences between the simulated Sahel and the rescaled NARI to estimate the simulated fast response to forcing, but this would imply a fast response to ALL in CMIP5 (Figure 9e) that is inconsistent with the uncoupled estimate (purple, Figure 5c): instead of wetting the Sahel, it consists of a drying response to increasing GHG of$-0.0042 \pm 0.0036 \ \frac{\mathrm{mm}}{day*year}$ (Figure 9h). Whether we should interpret this as a fast response or a non-NARI-mediated response to SST, this component of the forced response helps delay and increase the severity of the minimum in precipitation in ALL relative to the AA simulations.

The estimated fast responses for CMIP6 are displayed in Figure 10 in a fashion similar to Figure 2, and are compared to the fast response obtained as the difference between amip-hist and amip-piF simulations (purple, as in Figure 5c). Unlike the fast response in CMIP5, the ALL fast response in CMIP6 matches the AMIP fast response significantly better than noise (r = 0.51, sRMSE = 0.87), giving us confidence that the NARI teleconnection strength estimated from amip-piF is valid in CMIP6 coupled simulations. Like the amip-hist fast response, the ALL fast response in CMIP6 displays wetting after 1980 that is roughly equal to the sum (burgundy dashed curve) of the fast responses to AA (b, magenta) and GHG (d, green). The simulated fast wetting after 1980 in the ALL simulations (a, blue) is smaller than in the AMIP simulations, as expected if amip-hist is double-counting radiative forcing, but is still larger than our estimate of the optimal value (0.3 times the AMIP fast response), consistent with claims that the strength of radiative forcing is overestimated in the coupled simulations.

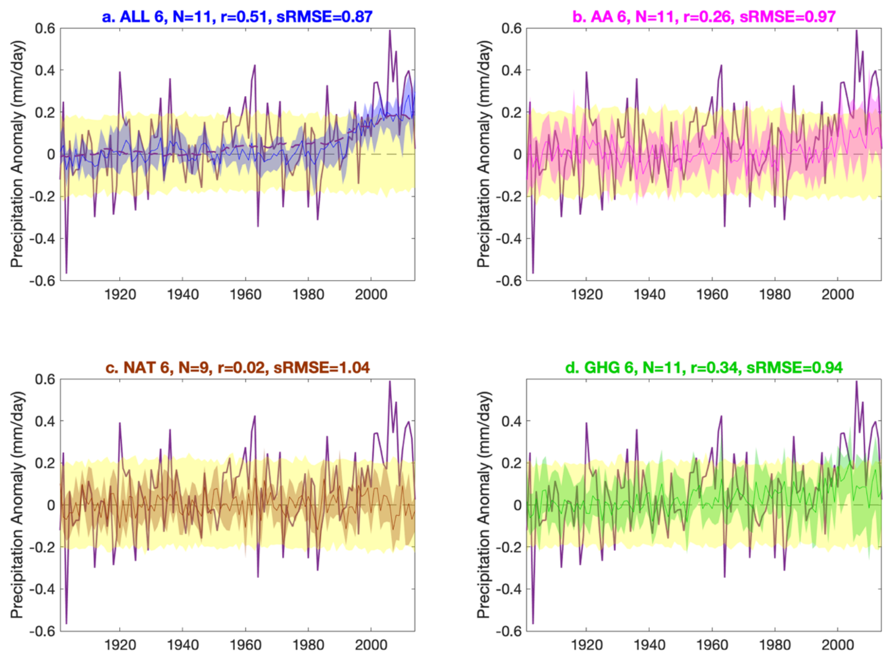

Fig. 10. Compares the fast Sahelian precipitation response to forcing in AMIP simulations (purple, as in Figure 5c) to the estimated fast component of the precipitation MMMs in coupled CMIP6 simulations (precipitation − 0.87*NARI; the difference between the colored and light blue curves in the left column of Figure 9) forced with ALL (a, blue), AA (b, magenta), NAT (c, brown), and GHG (d, green). Similar to Figure 2, the colored shaded areas denote the bootstrapping confidence interval of this difference, and the yellow shaded areas, which represent the magnitude of noise in the fast MMMs, are the confidence intervals of the MMM of randomized bootstrapped differences between precipitation and 0.87*NARI in piC simulations. Panel (a) additionally shows a 20-year running mean of the sum of the AA, NAT, and GHG fast MMMs (burgundy dashed curve). The label shows the number of institutions used for each CMIP6 MMM (N), the correlation of the fast MMM with the AMIP fast response (r), and the standardized root mean squared error of the CMIP6 MMM with observations (sRMSE).

Though NARI in the GHG simulations differs between CMIP5 and CMIP6, most of the difference in simulated forced precipitation between CMIP5 and CMIP6 is not mediated by a linear relationship with NARI, and can be attributed to the fact that the GHG- and AA-induced drying in CMIP5 is replaced with AA- and GHG-induced wetting in CMIP6. Whether the GHG-induced drying in CMIP5 is a fast response to forcing or a response mediated by SST in ocean basins other than the Atlantic cannot be firmly established by this analysis, but we offer our perspective below.

5. Discussion

Using SST (and specifically NARI) as a mediator, we have established that the failure of CMIP coupled models to simulate observed Sahel rainfall stems from their inability to simulate observed SST, especially

NA, and that the differences in simulation of Sahel rainfall between CMIP5 and CMIP6 stem from differences in mechanisms not mediated by a linear teleconnection with NARI. (Let's denote the difference between simulated precipitation and scaled NARI as $P_{nonNARI}$). We initially suggested that $P_{nonNARI}$ provides a good measure of the fast (non-SST-related) response to forcing because of the prominence of the NARI-Sahel teleconnection in observations and AMIP-style simulations of the 20th century. But without examining further mediators, we cannot decisively rule out the possibility that $P_{nonNARI}$ captures teleconnections with other ocean basins or nonlinearities in the NARI teleconnection. Which explanation is most likely?

The $P_{nonNARI}$ indices in CMIP5 and CMIP6 are nearly opposite. If we assume that both represent a fast response to forcing, we need to conclude that increasing GHG (or reducing AA) lead to fast wetting in CMIP6, but drying in CMIP5.

The interpretation of $P_{nonNARI}$ in CMIP6 as a fast response is more consistent with theory. First, increasing rainfall is consistent with theory linking reduced aerosol concentrations to fast surface warming and decreasing optical depth of the atmosphere (Allen and Ingram 2002; Rosenfeld et al. 2008), although a couple highly non-linear simulations suggest the fast precipitation response of the Sahel to changing AA in the 20th century was drying whether AA forcing was increasing or decreasing (Hirasawa et al. 2020). Second, it is generally accepted that the fast response of the Sahel to GHG is wetting (e.g. Biasutti 2013; Gaetani et al. 2017; Giannini 2010; Haarsma et al. 2005). The good match in the estimated fast response between coupled CMIP6 simulations and the amip-hist simulations increases our confidence that the deviations from the NARI-mediated slow response to forcing in CMIP6 really reflect a fast response to forcing. The same cannot be said for CMIP5.

We noted in Section 4.c that NARI only explains 36% of simulated SST-forced variability in the amip-piF simulations, leaving room for the influence of other ocean basins or SST indices on Sahel precipitation. Indeed, this is consistent with GK19: while they argue that NARI is the primary indicator for 20th century Sahel rainfall, they also argue that p1, which is approximately (NA+GT)/2 and is intended to capture the effects of uniform global warming, plays a secondary—but important—role in the 20th century and a dominant role in the future. In CMIP5, $P_{nonNARI}$may capture not the fast responses to forcing, but slow drying in response to uniform global warming, consistent with previous literature (e.g. Gaetani et al. 2017). In this read, the differences in simulation of Sahel rainfall between CMIP5 and CMIP6 are due to a combination of changes in the fast response to forcing and the influence of SST patterns not captured by NARI.

6. Summary and Conclusions

In this paper, we decompose simulated Sahelian precipitation into (1) teleconnections with SST, (2) fast, atmospheric- and land-mediated responses to forcing, (3) atmospheric noise, (4) forced SST variability, and (5) internal SST variability, in order to determine why the 5th and 6th generations of CMIP differ in their simulation of Sahel rainfall, and why both ensembles are inconsistent with observed Sahel precipitation variability.

CMIP6 atmospheric simulations forced with observed SST alone capture observed Sahel precipitation quite well (r=0.6), and, in combination with atmospheric white noise, are able to reproduce the power of observed low-frequency variability. This is a welcome improvement from previous generations of climate models. Including radiative forcing alongside observed SST barely changes simulated precipitation, suggesting that the fast response is small and plays a secondary role to SST-forced precipitation variability. We summarize the Sahel teleconnections with global SST as a linear relationship with an index of the warming of the North Atlantic relative to the global Tropics (NARI), which explains about 36% of the simulated precipitation response to observed SST. The simulated NARI teleconnection is measured as $0.87 \pm 0.26 \frac{\mathrm{mm}}{day*C}$, consistent with the strength of the observed teleconnection. We conclude that the observed SST history and simulated teleconnections in atmospheric simulations are together necessary and sufficient to capture the timing and magnitude of the low-frequency droughts and pluvials in 20thcentury Sahel rainfall.

In coupled simulations, the NARI-Sahel teleconnection is consistent with AMIP simulations, but NARI's variability – which mostly comes from North Atlantic SST (NA) – differs from the observed. In simulations, AA cause a cooling trend and GHG cause a warming trend with magnitudes comprable to the observed, but no combination of forcing agents produces a decadal-scale oscillation in NA in either CMIP5 or CMIP6, and only three CMIP6 models (out of 25 CMIP5 and 30 CMIP6 models) are able to generate internal SST variability commensurate to the residual (the difference between total and radiatively forced) low-frequency variability. How do we reconcile our results with those claiming that the observed Atlantic Multidecadal Variability (AMV) is externally forced (mainly by AA; Bellomo et al. 2018; Booth et al. 2012; Hirasawa et al. 2020; Hua et al. 2019; Murphy et al. 2017)? The discrepancy can be explained because these studies examine only one or two models (Booth et al. 2012; Hirasawa et al. 2020) or subtract a linear trend from simulated NA before comparing to observations (Bellomo et al. 2018; Hua et al. 2019; Murphy et al. 2017), thus inducing low-frequency variability in the simulated monotonic decreasing step function. Moreover, a prominent role for internal variability cannot yet be dismissed, as suggested by Yan et al. (2018), who, consistent with our analysis, find that most models do not capture observed AMOC variability. The NARI-mediated slow response to external radiative forcing is to dry the Sahel slightly in the 60s and to wet it immediately afterwards; this does not, in isolation, explain the timing or magnitude of the observed drought or recovery. Furthermore, forced NARI variability is small in the first half of the century. We are led to conclude that either the pattern of the simulated SST response to forcing in coupled models is incorrect or the Sahelian precipitation response to internal SST variability overshadowed the response to external radiative forcing in the 20$^{\text{th}}$century, at least up to the mid-1960s.

While we can ascribe the deficiency of 20$^{\text{th}}$ century Sahel rainfall simulations in both CMIP5 and CMIP6 coupled models to their simulations of SST, NARI is not the main explanation for the differences in forced Sahel rainfall between the two ensembles, since it is quite similar in CMIP5 and CMIP6 ALL simulations. The difference, rather, is in P_{nonNARI}: the component of Sahel rainfall that comes either from the influence of other SST patterns or from the fast response to forcing. CMIP6 underperforms relative to CMIP5 because P_{nonNARI} includes substantial fast wetting responses to increasing GHG and decreasing AA, comparable in magnitude to the NARI-related component. In contrast, P_{nonNARI} in CMIP5 is drying, likely in response to uniform SST warming. Sahel drying in response to uniform warming is strong in models that simulate a deeper ascent profile, but weak otherwise (Hill et al 2017), so it is possible that newer parameterizations and higher resolution have changed the sensitivity to this forcing in the latest generation of models.

This work has shown that, while there has been progress in the simulation of the Sahel's response to global SST, much remains uncertain in the simulation of the pathways of Sahel multi-decadal variability, especially in the amplitude and timing of forced and natural SST anomalies in the Atlantic and in the fast and slow response of rainfall to GHG forcing. Differing mechanisms can lead to similar time evolutions in observations and simulations; to avoid this pitfall, future work should focus on evaluating in more detail the hypothesized pathways of the Sahel response to anthropogenic emissions and oceanic internal variability in order to further categorize model performance and improve predictions of the future.

Acknowledgments.

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 for producing and making available their model output (listed in Tables S1-S3 of this paper). 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. The CMIP6 data used in this study are available in Google cloud storage (https://console.cloud.google.com/storage/browser/cmip6) as a result of a grant to the Pangeo project (https://pangeo.io/). We thank Haibo Liu for preparing the CMIP5 data for use, and Naomi Henderson for transferring needed CMIP6 simulations to the cloud, and for aiding in data access and general technical support throughout the project. We additionally thank Alessandra Giannini for her guidance throughout the project. This research was supported by the U.S. National Science Foundation Grant AGS-1612904.

Data Availability Statement.

Observational data from the Global Precipitation Climatology Center (GPCC, Becker et al. 2013) and the National Oceanic and Atmospheric Administration's (NOAA) Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5, Huang et al. 2017) are freely available online (see https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html and https://www.ncei.noaa.gov/products/extended-reconstructed-sst, respectively). CMIP5 (CMIP5, Taylor et al. 2012) and CMIP6 (Eyring et al. 2016) model data is freely available through the Earth System Grid (see https://esgf-node.llnl.gov/projects/esgf-llnl/).

REFERENCES

Ackerley, D., B. B. B. Booth, S. H. E. Knight, E. J. Highwood, D. J. Frame, M. R. Allen, and D. P. Rowell, 2011: Sensitivity of twentieth-century Sahel rainfall to sulfate aerosol and CO2 forcing. *J. Climate* , **24,** 4999-5014, https://doi.org/10.1175/JCLI-D-11-00019.1.

Adam, O., T. Bischoff, and T. Schneider, 2016: Seasonal and interannual variations of the energy flux equator and ITCZ. Part II: Zonally varying shifts of the ITCZ. *J. Climate* , **29,** 7281-7293, https://doi.org/10.1175/JCLI-D-15-0710.1.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle.*Nature* , **419,** 228-232, https://doi.org/10.1038/nature01092.

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

Bellomo, K., L. N. Murphy, M. A. Cane, A. C. Clement, and L. M. Polvani, 2018: Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dynam.* , **50,** 3687-3698, https://doi.org/10.1007/s00382-017-3834-3.

Biasutti, M., 2013: Forced Sahel rainfall trends in the CMIP5 archive. *J. Geophys. Res.-Atmos.* ,**118,** 1613-1623, https://doi.org/10.1002/jgrd.50206.

——, 2019: Rainfall trends in the African Sahel: Characteristics, processes, and causes. *Wiley Interdisciplinary Reviews: Climate Change* , **10** , https://doi.org/10.1002/wcc.591.

Biasutti, M., and A. Giannini, 2006: Robust Sahel drying in response to late 20th century forcings.*Geophys. Res. Lett.* , **33** , https://doi.org/10.1029/2006GL026067.

Biasutti, M., I. M. Held, A. H. Sobel, and A. Giannini, 2008: SST forcings and Sahel rainfall variability in simulations of the twentieth and twenty-first centuries. *J. Climate* , **21,** 3471-3486, https://doi.org/10.1175/2007JCLI1896.1.

Biasutti, M., and Coauthors, 2018: Global energetics and local physics as drivers of past, present and future monsoons. *Nature Geoscience* , **11,** 392-400, https://doi.org/10.1038/s41561-018-0137-1.

Bonfils, C. J., B. D. Santer, J. C. Fyfe, K. Marvel, T. J. Phillips, and S. R. Zimmerman, 2020: Human influence on joint changes in temperature, rainfall and continental aridity. *Nat. Clim. Change* , **10,** 726-731, https://doi.org/10.1038/s41558-020-0821-1.

Booth, B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability.*Nature* , **484,** 228-232, https://doi.org/10.1038/nature10946.

Camberlin, P., S. Janicot, and I. Poccard, 2001: Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *International Journal of Climatology: A Journal of the Royal Meteorological Society* ,**21,** 973-1005, https://doi.org/10.1002/joc.673.

Caminade, C., and L. Terray, 2010: Twentieth century Sahel rainfall variability as simulated by the ARPEGE AGCM, and future changes. *Climate Dynam.* , **35,** 75-94, https://doi.org/10.1007/s00382-009-0545-4.

Chang, C.-Y., J. Chiang, M. Wehner, A. Friedman, and R. Ruedy, 2011: Sulfate aerosol control of tropical Atlantic climate over the twentieth century. *J. Climate* ,**24,** 2540-2555, https://doi.org/10.1175/2010JCLI4065.1.

Dong, B., and R. Sutton, 2015: Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall. *Nat. Clim. Change* , **5,** 757-761, https://doi.org/10.1038/nclimate2664.

Donohoe, A., J. Marshall, D. Ferreira, and D. Mcgee, 2013: The relationship between ITCZ location and cross-equatorial atmospheric heat transport: From the seasonal cycle to the Last Glacial Maximum. *J. Climate* , **26,** 3597-3618, https://doi.org/10.1175/JCLI-D-12-00467.1.

Eade, R., D. Stephenson, A. Scaife, and D. Smith, 2021: Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic Oscillation? *Climate Dynam.* **,** 1-14, https://doi.org/10.1007/s00382-021-05978-4.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* ,**9,** 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature* , **320,** 602-607, https://doi.org/10.1038/320602a0.

Gaetani, M., and Coauthors, 2017: West African monsoon dynamics and precipitation: the competition between global SST warming and CO 2 increase in CMIP5 idealized simulations.*Climate Dynam.* , **48,** 1353-1373, https://doi.org/10.1007/s00382-016-3146-z.

Giannini, A., 2010: Mechanisms of climate change in the semiarid African Sahel: The local view. *J. Climate* , **23,** 743-756, https://doi.org/10.1175/2009JCLI3123.1.

Giannini, A., and A. Kaplan, 2019: The role of aerosols and greenhouse gases in Sahel drought and recovery.*Climatic Change* , **152,** 449-466, https://doi.org/10.1007/s10584-018-2341-9.

Giannini, A., R. Saravanan, and P. Chang, 2003: Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science* , **302,** 1027-1030, https://doi.org/10.1126/science.1089357.

Giannini, A., M. Biasutti, I. M. Held, and A. H. Sobel, 2008: A global perspective on African climate.*Climatic Change* , **90,** 359-383, https://doi.org/10.1007/s10584-008-9396-y.

Giannini, A., S. Salack, T. Lodoun, A. Ali, A. Gaye, and O. Ndiaye, 2013: A unifying view of climate change in the Sahel linking intra-seasonal, interannual and longer time scales.*Environ. Res. Lett.* , **8,** 024010, https://doi.org/10.1088/1748-9326/8/2/024010.

Haarsma, R. J., F. M. Selten, S. L. Weber, and M. Kliphuis, 2005: Sahel rainfall variability and response to greenhouse warming. *Geophys. Res. Lett.* , **32** , https://doi.org/10.1029/2005GL023232.

Han, Z., F. Luo, S. Li, Y. Gao, T. Furevik, and L. Svendsen, 2016: Simulation by CMIP5 models of the Atlantic multidecadal oscillation and its climate impacts.*Advances in Atmospheric Sciences* , **33,** 1329-1342, https://doi.org/10.1007/s00376-016-5270-4.

Haywood, J. M., A. Jones, N. Bellouin, and D. Stephenson, 2013: Asymmetric forcing from stratospheric aerosols impacts Sahelian rainfall. *Nat. Clim. Change* ,**3,** 660-665, https://doi.org/10.1038/nclimate1857.

Held, I. M., T. L. Delworth, J. Lu, K. u. Findell, and T. Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centuries. *Proc. Nat. Acad. Sci.* , **102,**17891-17896, https://doi.org/10.1073/pnas.0509057102.

Herman, R. J., A. Giannini, M. Biasutti, and Y. Kushnir, 2020: The effects of anthropogenic and volcanic aerosols and greenhouse gases on twentieth century Sahel precipitation. *Scientific reports* , **10,** 1-11, https://doi.org/10.1038/s41598-020-68356-w.

Hill, S. A., 2019: Theories for past and future monsoon rainfall changes. *Current Climate Change Reports* , **5,** 160-171, https://doi.org/10.1007/s40641-019-00137-8.

Hirasawa, H., P. J. Kushner, M. Sigmond, J. Fyfe, and C. Deser, 2020: Anthropogenic Aerosols Dominate Forced Multidecadal Sahel Precipitation Change through Distinct Atmospheric and Oceanic Drivers. *J. Climate* , **33,**10187-10204, https://doi.org/10.1175/JCLI-D-19-0829.1.

Hoerling, M., J. Hurrell, J. Eischeid, and A. Phillips, 2006: Detection and attribution of twentieth-century northern and southern African rainfall change.*J. Climate* , **19,** 3989-4008, https://doi.org/10.1175/JCLI3842.1.

Hua, W., A. Dai, L. Zhou, M. Qin, and H. Chen, 2019: An Externally Forced Decadal Rainfall Seesaw Pattern Over the Sahel and Southeast Amazon. *Geophys. Res. Lett.* , **46,**923-932, https://doi.org/10.1029/2018GL081406.

Huang, B., and Coauthors, 2017: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Climate* ,**30,** 8179-8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Hwang, Y. T., D. M. Frierson, and S. M. Kang, 2013: Anthropogenic sulfate aerosol and the southward shift of tropical precipitation in the late 20th century. *Geophys. Res. Lett.* , **40,** 2845-2850.

Iles, C. E., and G. C. Hegerl, 2014: The global precipitation response to volcanic eruptions in the CMIP5 models. *Environ. Res. Lett.* , **9** , https://doi.org/10.1088/1748-9326/9/10/104012.

Kang, S. M., D. M. Frierson, and I. M. Held, 2009: The tropical response to extratropical thermal forcing in an idealized GCM: The importance of radiative feedbacks and convective parameterization. *J. Atmos. Sci.* , **66,** 2812-2827, https://doi.org/10.1175/2009JAS2924.1.

Kang, S. M., I. M. Held, D. M. Frierson, and M. Zhao, 2008: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *J. Climate* , **21,** 3521-3532, https://doi.org/10.1175/2007JCLI2146.1.

Kawase, H., M. Abe, Y. Yamada, T. Takemura, T. Yokohata, and T. Nozawa, 2010: Physical mechanism of long-term drying trend over tropical North Africa. *Geophys. Res. Lett.* , **37** , https://doi.org/10.1029/2010GL043038.

Klimont, Z., S. J. Smith, and J. Cofala, 2013: The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. *Environ. Res. Lett.* , **8** , https://doi.org/10.1088/1748-9326/8/1/014003.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.* , **33** , https://doi.org/10.1029/2006GL026242.

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate.*Geophys. Res. Lett.* , **32** , https://doi.org/10.1029/2005GL024233.

Knutson, T. R., and S. Manabe, 1995: Time-mean response over the tropical Pacific to increased C02 in a coupled ocean-atmosphere model. *J. Climate* , **8,**2181-2199, https://doi.org/10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2.

Kucharski, F., N. Zeng, and E. Kalnay, 2013: A further assessment of vegetation feedback on decadal Sahel rainfall variability. *Climate Dynam.* , **40,**1453-1466, DOI 10.1007/s00382-012-1397-x.

Losada, T., B. Rodriguez-Fonseca, E. Mohino, J. Bader, S. Janicot, and C. R. Mechoso, 2012: Tropical SST and Sahel rainfall: A non-stationary relationship. *Geophys. Res. Lett.* , **39** , https://doi.org/10.1029/2012GL052423.

Lu, J., 2009: The dynamics of the Indian Ocean sea surface temperature forcing of Sahel drought. *Climate Dynam.* , **33,** 445-460, https://doi.org/10.1007/s00382-009-0596-6.

Marvel, K., M. Biasutti, and C. Bonfils, 2020: Fingerprints of external forcings on Sahel rainfall: aerosols, greenhouse gases, and model-observation discrepancies. *Environ. Res. Lett.* , **15** , https://doi.org/10.1088/1748-9326/ab858e.

Menary, M. B., and Coauthors, 2020: Aerosol-forced AMOC changes in CMIP6 historical simulations. *Geophys. Res. Lett.* , **47** , https://doi.org/10.1029/2020GL088166.

Murphy, L. N., K. Bellomo, M. Cane, and A. Clement, 2017: The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophys. Res. Lett.* ,**44,** 2472-2480, https://doi.org/10.1002/2016GL071337.

Neelin, J., C. Chou, and H. Su, 2003: Tropical drought regions in global warming and El Nino teleconnections. *Geophys. Res. Lett.* , **30** , https://doi.org/10.1029/2003GL018625.

Okonkwo, C., and Coauthors, 2015: Combined effect of El Niño southern oscillation and Atlantic multidecadal oscillation on Lake Chad level variability. *Cogent Geoscience* , **1** , https://doi.org/10.1080/23312041.2015.1117829.

Palmer, T., 1986: Influence of the Atlantic, Pacific and Indian oceans on Sahel rainfall. *Nature* ,**322,** 251-253, https://doi.org/10.1038/322251a0.

Parhi, P., A. Giannini, P. Gentine, and U. Lall, 2016: Resolving contrasting regional rainfall responses to El Niño over tropical Africa. *J. Climate* , **29,** 1461-1476, https://doi.org/10.1175/JCLI-D-15-0071.1.

Park, J.-y., J. Bader, and D. Matei, 2016: Anthropogenic Mediterranean warming essential driver for present and future Sahel rainfall. *Nat. Clim. Change* , **6,**941-945, https://doi.org/10.1038/nclimate3065.

Pearl, J., M. Glymour, and N. P. Jewell, 2016: *Causal inference in statistics: A primer.* John Wiley & Sons, 136 pp.

Polson, D., M. Bollasina, G. Hegerl, and L. Wilcox, 2014: Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols. *Geophys. Res. Lett.* ,**41,** 6023-6029, https://doi.org/10.1002/2014GL060811.

Pomposi, C., Y. Kushnir, and A. Giannini, 2015: Moisture budget analysis of SST-driven decadal Sahel precipitation variability in the twentieth century. *Climate Dynam.* , **44,** 3303-3321, https://doi.org/10.1007/s00382-014-2382-3.

Pomposi, C., A. Giannini, Y. Kushnir, and D. E. Lee, 2016: Understanding Pacific Ocean influence on interannual precipitation variability in the Sahel. *Geophys. Res. Lett.* , **43,** 9234-9242, https://doi.org/10.1002/2016GL069980.

Qin, M., A. Dai, and W. Hua, 2020: Quantifying contributions of internal variability and external forcing to Atlantic multidecadal variability since 1870. *Geophys. Res. Lett.* , **47** , https://doi.org/10.1029/2020GL089504.

Rahmstorf, S., J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, and E. J. Schaffernicht, 2015: Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat. Clim. Change* , **5,** 475-480, https://doi.org/10.1038/nclimate2554.

Rodríguez-Fonseca, B., and Coauthors, 2015: Variability and predictability of West African droughts: A review on the role of sea surface temperature anomalies. *J. Climate* ,**28,** 4034-4060, https://doi.org/10.1175/JCLI-D-14-00130.1.

Rosenfeld, D., and Coauthors, 2008: Flood or drought: How do aerosols affect precipitation? *Science* ,**321,** 1309-1313, https://doi.org/10.1126/science.1160606.

Rotstayn, L. D., and U. Lohmann, 2002: Tropical rainfall trends and the indirect aerosol effect. *J. Climate* , **15,** 2103-2116, https://doi.org/10.1175/1520-0442(2002)015<2103:TRTATI>2.0.CO;2.

Scaife, A., and Coauthors, 2009: The CLIVAR C20C project: selected twentieth century climate events.*Climate Dynam.* , **33,** 603-614, https://doi.org/10.1007/s00382-008-0451-1.

Schneider, T., T. Bischoff, and G. H. Haug, 2014: Migrations and dynamics of the intertropical convergence zone. *Nature* , **513,** 45-53, https://doi.org/10.1038/nature13636.

Smith, S. J., J. v. Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias, 2011: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.* , **11,** 1101-1116, https://doi.org/10.5194/acp-11-1101-2011.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature.*J. Climate* , **15,** 2702-2706, https://doi.org/10.1175/1520-0442(2002)015<2702:TESITT>2.0.CO;2.

Sutton, R. T., and D. L. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science* , **309,** 115-118, https://doi.org/10.1126/science.1109496.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design.*Bull. Am. Meteorol. Soc.* , **93,** 485-498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate* , **22,** 1469-1481, https://doi.org/10.1175/2008JCLI2561.1.

Undorf, S., D. Polson, M. Bollasina, Y. Ming, A. Schurer, and G. Hegerl, 2018: Detectable impact of local and remote anthropogenic aerosols on the 20th century changes of West African and South Asian monsoon precipitation. *J. Geophys. Res.-Atmos.* , **123,** 4871-4889, https://doi.org/10.1029/2017JD027711.

Vellinga, M., and Coauthors, 2016: Sahel decadal rainfall variability and the role of model horizontal resolution. *Geophys. Res. Lett.* , **43,** 326-333, https://doi.org/10.1002/2015GL066690.

Westervelt, D., and Coauthors, 2017: Multimodel precipitation responses to removal of US sulfur dioxide emissions. *J. Geophys. Res.-Atmos.* , **122,** 5024-5038, https://doi.org/10.1002/2017JD026756.

Yan, X., R. Zhang, and T. R. Knutson, 2018: Underestimated AMOC Variability and Implications for AMV and Predictability in CMIP Models. *Geophys. Res. Lett.* , **45,**4319-4328, https://doi.org/10.1029/2018GL077378.

——, 2019: A multivariate AMV index and associated discrepancies between observed and CMIP5 externally forced AMV. *Geophys. Res. Lett.* , **46,** 4421-4431, https://doi.org/10.1029/2019GL082787.

Zhang, R., 2017: On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. *Geophys. Res. Lett.* , **44,** 7865-7875, https://doi.org/10.1002/2017GL074342.

Zhang, R., and T. L. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.* , **33** , https://doi.org/10.1029/2006GL026267.

Zhang, R., R. Sutton, G. Danabasoglu, T. L. Delworth, W. M. Kim, J. Robson, and S. G. Yeager, 2016: Comment on "The Atlantic Multidecadal Oscillation without a role for ocean circulation". *Science* , **352,** 1527-1527, https://doi.org/10.1126/science.aaf1660.

Zhang, R., and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.* , **70,** 1135-1144, https://doi.org/10.1175/JAS-D-12-0331.1.

Hosted file

orals paper submit supplement.docx available at https://authorea.com/users/529345/articles/599116-drivers-of-low-frequency-sahel-precipitation-variability-comparing-cmip5-and-cmip6-with-observations

Drivers of Low-Frequency Sahel Precipitation Variability: Comparing CMIP5 and CMIP6 with Observations

Rebecca Jean Herman,[a] Michela Biasutti,[b] Yochanan Kushnir[b]

[a] *Department of Earth and Environmental Sciences of Columbia University, New York, NY*

[b] *Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY*

Corresponding author: Rebecca Herman, rebecca.herman@columbia.edu

ABSTRACT

We examine and contrast the simulation of Sahel rainfall in phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6). On average, both ensembles grossly underestimate the magnitude of low-frequency variability in Sahel rainfall. But while CMIP5 partially matches the timing and pattern of observed multi-decadal rainfall swings in its historical simulations, CMIP6 does not. To classify model deficiency, we use the previously-established link between changes in Sahelian precipitation and the North Atlantic Relative Index (NARI) for sea surface temperature (SST) to partition all influences on Sahelian precipitation into five components: (1) teleconnections to SST variations; the effects of (2) atmospheric and (3) SST variability internal to the climate system; (4) the SST response to external radiative forcing; and (5) the "fast" response to forcing, which is not mediated by SST. CMIP6 atmosphere-only simulations indicate that the fast response to forcing plays only a small role relative to the predominant effect of observed SST variability on low-frequency Sahel precipitation variability, and that the strength of the NARI teleconnection is consistent with observations. Applying the lessons of atmosphere-only models to coupled settings, we imply that the failure of coupled models in simulating 20th century Sahel rainfall derives from their failure to simulate the observed combination of forced and internal variability in SST. Yet differences between CMIP5 and CMIP6 Sahel precipitation do not mainly derive from differences in NARI, but from either their fast response to forcing or the role of other SST patterns.

1. Introduction

The semi-arid region bordering the North African Savanna and the Sahara Desert, known as the Sahel, received much scientific attention since it experienced unparalleled dramatic rainfall variability in the second half of the 20th century. The importance of teleconnections between Sahel precipitation and global sea surface temperature (SST) was demonstrated in the early stages of Sahel climate variability research (Folland et al. 1986; Giannini et al. 2003; Knight et al. 2006; Palmer 1986; Zhang and Delworth 2006), and has been further reinforced in more recent studies (Okonkwo et al. 2015; Parhi et al. 2016; Park et al. 2016; Pomposi et al. 2015; Pomposi et al. 2016; Rodríguez-Fonseca et al. 2015 and references therein). But while the dominant role of SST in driving the pacing (though not necessarily the full magnitude) of 20th century Sahel rainfall variability is unquestioned (Biasutti 2019), there is still debate on whether the evolution of SST and the related Sahel precipitation variability were externally forced (Ackerley et al. 2011; Biasutti 2013; Biasutti and Giannini 2006; Biasutti et al. 2008; Bonfils et al. 2020; Dong and Sutton 2015; Giannini and Kaplan 2019; Haarsma et al. 2005; Haywood et al. 2013; Held et al. 2005; Hirasawa et al. 2020; Hua et al. 2019; Iles and Hegerl 2014; Kawase et al. 2010; Marvel et al. 2020; Polson et al. 2014; Undorf et al. 2018; Westervelt et al. 2017) or the manifestation of variability internal to the climate system (IV, Sutton and Hodson 2005; Ting et al. 2009; Zhang and Delworth 2006).

Recently, Herman et al. (2020, hereafter H20) investigated multi-model means (MMM) of historical simulations from the Coupled Model Intercomparison Project phase 5 (CMIP5, Taylor et al. 2012), and found that anthropogenic aerosols (AA) and volcanic aerosols (VA), but not greenhouse gases (GHG), were responsible for forcing simulated Sahelian precipitation that correlates well with observations, with AA alone responsible for the low-frequency component of simulated variability. This conclusion appeared consistent with previous claims that AA emissions, which increased until the 1970s and then decreased in response to clean air initiatives (Klimont et al. 2013; Smith et al. 2011), caused multi-decadal variability in Sahel precipitation via changes in Northern Hemisphere surface temperature (Ackerley et al. 2011; Haywood et al. 2013; Hwang et al. 2013; Undorf et al. 2018), or specifically via multidecadal variability in North Atlantic SST (the Atlantic Multidecadal Variability, AMV; Booth et al. 2012; Hua et al. 2019). However, H20 also found that the simulated rainfall response to forcing has little low-frequency power relative to observations, and that simulated IV is unable to account for this difference.

H20 and most other attribution studies do not examine in depth the pathways through which AA (and for that matter, IV and other external forcing agents) affect Sahel precipitation. Thus, H20 did not determine whether the discrepancy between CMIP5 simulations and observations represents an underestimate of aerosol indirect effects and climate feedbacks that amplify the simulated precipitation response to AA, or a fundamental inability of the models to simulate aspects of the observed climate response to forcing or observed modes of IV. Identifying the deficiencies in model representation of the pathways by which external forcing and IV influence the West African Monsoon and Sahel rainfall is essential for attribution of 20th century changes and also for prediction of this region’s climate future, as model simulations don’t even agree on the sign of future precipitation changes in the Sahel (Biasutti 2013).

Here, we use the well-established link between SST and Sahel precipitation to decompose the effects of individual external forcing agents (F) and internal variability (IV) on Sahel precipitation (P) into five path components, presented in Figure 1: (1) teleconnections that communicate variations in SST to variations in P (indicated by the arrow $\vec{t}$); (2) the "fast" atmospheric and land-mediated effect of external forcing (F) on P ($\vec{f}$); (3) the direct effect of atmospheric IV on P ($\vec{a}$); (4) the effect of F on SST ($\vec{s}$); and (5) the impact of IV in the coupled climate system on SST ($\vec{o}$). The path F → SST → P is the "slow," SST-mediated effect of F on P.

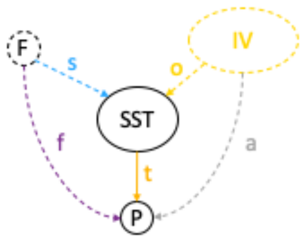

Fig. 1. Causal diagram relating external forcings (F), internal variability (IV), sea surface temperatures (SST), and Sahelian precipitation (P) via directional causal arrows. Unobserved variables and their causal effects are presented with dashed lines, while observed variables are presented with solid lines.

Characterization of these path components has been controversial. Firstly, separating the SST response to forcing ($\vec{s}$) from SST variability internal to the climate system ($\vec{o}$) has proven difficult (top of diagram). In particular, there is significant debate over whether observed AMV is a response to external forcing (Booth et al. 2012; Chang et al. 2011; Hua et al. 2019; Menary et al. 2020; Rotstayn and Lohmann 2002) or mainly an expression of IV in the Atlantic Meridional Overturning Circulation (AMOC, Han et al. 2016; Knight et al. 2005; Qin et al. 2020; Rahmstorf et al. 2015; Sutton and Hodson 2005; Ting et al. 2009; Yan et al. 2019; Zhang 2017; Zhang et al. 2016; Zhang et al. 2013) that is underestimated in models (Yan et al. 2018). This debate has been hard to resolve partially because IV in AMOC and aerosol forcing may have coincided by chance in the 20th century (Qin et al. 2020). Next, examine the bottom of the diagram. The effect of the observed SST field on Sahel precipitation ($\vec{t}$) can be directly estimated using atmosphere-only simulations, but while these simulations capture the pattern of observed Sahel precipitation variability, many fail to capture its full magnitude (Biasutti 2019; e.g. Hoerling et al. 2006; Scaife et al. 2009). This could reflect an underestimate in climate models of the strength of SST teleconnections, which could be resolution dependent (Vellinga et al. 2016), or of land-climate feedbacks that amplify the teleconnections ($\vec{t}$), such as vegetation changes (Kucharski et al. 2013). But it could also reflect a significant additional role in the observations for a fast response to forcing ($\vec{f}$) that confounds the SST-forced signal [P ← F → SST → P; see Pearl et al. (2016) for notation] or coincides with it by chance.

To examine the path components in coupled simulations, we need a parsimonious characterization of the relationship between SST and Sahel precipitation. Giannini et al. (2013) and Giannini and Kaplan (2019, hereafter GK19) identify the North Atlantic Relative Index (NARI), defined as the difference between average SST in the North Atlantic (NA) and in the Global Tropics (GT), as the dominant SST indicator of 20th century Sahel rainfall in observations and CMIP5 simulations. There are two main theories relating NARI to Sahelian precipitation (see Biasutti 2019; Hill 2019 for reviews of competing theories of monsoon rainfall changes). In the first, the "local view" (Giannini 2010), warming of GT causes even stronger warming throughout the tropical upper troposphere (Knutson and Manabe 1995; Parhi et al. 2016; Sobel et al. 2002), increasing thermodynamic stability across the tropics and inhibiting convection in an "upped ante" (Giannini and Kaplan 2019; Neelin et al. 2003) or "tropospheric stabilization" (Giannini et al. 2008; Lu 2009) mechanism. Warming of NA, on the other hand, is expected to thermodynamically increase moisture supply to the Sahel by increasing specific humidity over the NA, and thus destabilize the atmospheric column from the bottom up (GK19). The second theory interprets the relationship of Sahel precipitation to NARI, or, similarly, to the Atlantic meridional temperature gradient or the Interhemispheric Temperature Difference (ITD), as the result of an energetically-driven shift in the Intertropical Convergence Zone (ITCZ, Donohoe et al. 2013; Kang et al. 2009; Kang et al. 2008; Knight et al. 2006; Schneider et al. 2014) and the African rainbelt (e.g. Adam et al. 2016; Biasutti et al. 2018; Camberlin et al. 2001; Caminade and Terray 2010; Hoerling et al. 2006; Hua et al. 2019; Pomposi et al. 2015; Westervelt et al. 2017). According to both theories, an increase in NARI should wet the Sahel while a decrease causes drying. Given the prominence of the NARI teleconnection in the 20th century and the assumption of linearity, we approximate the full slow response as the product of the NARI response to external forcing and the strength of the NARI-Sahel teleconnection.

This paper is organized as follows: Section 2 provides details on the simulations and observational data used in this analysis while Section 3 discusses the methods. In Section 4.a, we update H20's analysis to the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al. 2016), examining the total response to forcing (all paths from F to P) and internal variability (all paths from IV to P). We then evaluate the performance of the CMIP6 AMIP simulations, decomposing them into the path components from the bottom half of Figure 1 ($\vec{t}$, $\vec{f}$, and $\vec{a}$) in Section 4.b, and focusing on the NARI teleconnection in Section 4.c. Section 4.d decomposes coupled simulations of NARI into the path components from the top half of Figure 1 ($\vec{s}$ and $\vec{o}$), while Section 4.e evaluates the consistency of the NARI teleconnection established in Section 4.c with coupled simulations. Finally, in Section 4.f, we use simulated NARI and the simulated NARI teleconnection to decompose the total response of Sahel precipitation to external forcing in coupled simulations (examined in Section 4.a) into fast and slow components. We discuss how our results fit in with the existing literature in Section 5 before concluding in Section 6.

2. Data

We examine coupled "historical" simulations from CMIP5 (Taylor et al. 2012) and CMIP6 (Eyring et al. 2016) forced with four sets of forcing agents—AA alone, natural forcing alone (NAT, which includes VA as well as solar and orbital forcings), GHG alone, and all three simultaneously (ALL)—as well as pre-Industrial control (piC) simulations, in which all external forcing agents are held constant at pre-Industrial levels. We additionally examine CMIP6 amip-piForcing (amip-piF) simulations, in which atmospheric models are forced solely with observed SST, and CMIP6 amip-hist simulations, which are forced with observed SST and historical ALL radiative forcing. Calculations with CMIP5 utilize the period between 1901 and 2003 while calculations with CMIP6 extend to 2014.

In H20, we used all available institutions for each forcing subset. Here, in order to provide a more stringent comparison of the effects of different forcing agents, we exclude institutions from the coupled ensemble that do not provide AA, GHG, and ALL simulations, and from the AMIP ensemble if they do not provide both amip-piForcing and amip-hist simulations. We additionally exclude piC simulations that are shorter than the historical simulations as well as any simulations with data quality issues. Tables S1-S3 enumerate the simulations used in this analysis.

Precipitation observations are from the Global Precipitation Climatology Center (GPCC, Becker et al. 2013) version2018, and SST observations are from the National Oceanic and

Atmospheric Administration's (NOAA) Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5, Huang et al. 2017).

We analyze precipitation over the Sahel (12°-18°N and 20°W-40°E) and the SST indices of GK19: the North Atlantic (NA, 10°-40°N and 75°-15°W), the Global Tropics (GT, ocean surface in the latitude band 20°S-20°N), and the North Atlantic Relative index (NARI, the difference between NA and GT). All indices are spatially- and seasonally-averaged for July-September (JAS).

3. Methods

The multi-model mean (MMM) for a set of simulations consists of a 3-tiered weighted average over (1) individual simulations (runs) from each model, (2) models from each research institution, and (3) institutions in that ensemble. Details of the weighting are provided in H20; the results are robust to differences in weighting. Time series are not detrended, and anomalies are calculated relative to the period 1901-1950.

To evaluate the performance of the simulations relative to observations, we compute correlations (r), which capture similarity in frequency and phase, and root mean squared errors standardized by observed variance (sRMSE), which measure yearly differences in magnitude between the simulations and observations. An sRMSE of 0 represents a perfect match between simulations and observations, and 1 would result from comparing the observations with a constant time series.

To estimate uncertainty in the forced MMMs and associated metrics, we apply a bootstrapping technique to the last tier of the MMM as described in H20, yielding a probability distribution function (pdf) about the MMM and each metric. Due to the finite number of simulations, these pdfs underestimate the true magnitude of the uncertainty. We evaluate significance by applying a randomized bootstrapping technique, which increases the effective sample size, to the piC simulations with one significant improvement over H20: instead of using just one subset of each piC simulation at a random offset in the first tier of the MMM in each bootstrapping iteration, we take enough subsets to match the number of that model's historical runs. Done this way, the confidence intervals calculated using piC simulations accurately represent noise in the forced MMMs. PiC pdfs from the same ensemble differ slightly because many institutions provide a different number of simulations for different subsets of forcing agents (see Table S2). Where the piC pdfs and confidence intervals are similar enough, they are presented together with a single grey dotted curve and dashed line; when they differ, they are presented in the colors associated with the relevant forcings.

We perform a residual consistency test, which compares the power spectra (PS) of individual simulations to that of observations, with one significant modification over H20: we calculate the PS using the multi-taper method. Confidence intervals for the PS for observations and MMMs are given by the multi-taper method, without accounting for the uncertainty in the MMMs themselves. Mean PS by model are colored by climatological rainfall bias given by those simulations. The multi-model mean of these PS, or the "tiered mean", is calculated using the three tiers from the definition of the MMM, but without weights, since spectral power is not attenuated when averaging PS.

4. Results

a. Changes in CMIP6: Total Precipitation Response to Forcing and Internal Variability

If Sahelian precipitation is a linear combination of IV in the coupled climate system and variability forced by external agents, then the MMM over coupled simulations with differing initial conditions filters out atmospheric and oceanic IV ($\vec{a}$ and $\vec{o}$), leaving the fast and slow precipitation responses to external radiative forcing ($\vec{f}$ and F $\rightarrow$ SST $\rightarrow$ P). Figure 2 compares observed Sahelian precipitation anomalies (black, left ordinates) to the MMM anomalies of simulated Sahelian precipitation (right, amplified colored ordinates) in CMIP5 (dotted curves) and CMIP6 (solid curves) for four sets of forcing agents: ALL (a, blue), AA (b, magenta), natural forcing (c, "NAT," brown and red), and GHG (d, green). The figure also presents the bootstrapping 95% confidence intervals of the forced CMIP6 MMMs (blue, magenta, brown, and green shaded areas) and of MMMs over the CMIP6 piC simulations (yellow shaded areas) on the right ordinates. The width of the yellow shaded areas represents the magnitude of noise deriving from coincident IV in the MMMs. Differences in its width between panels arise from varying numbers of simulations for the different forcing subsets (see Methods and Table S2).

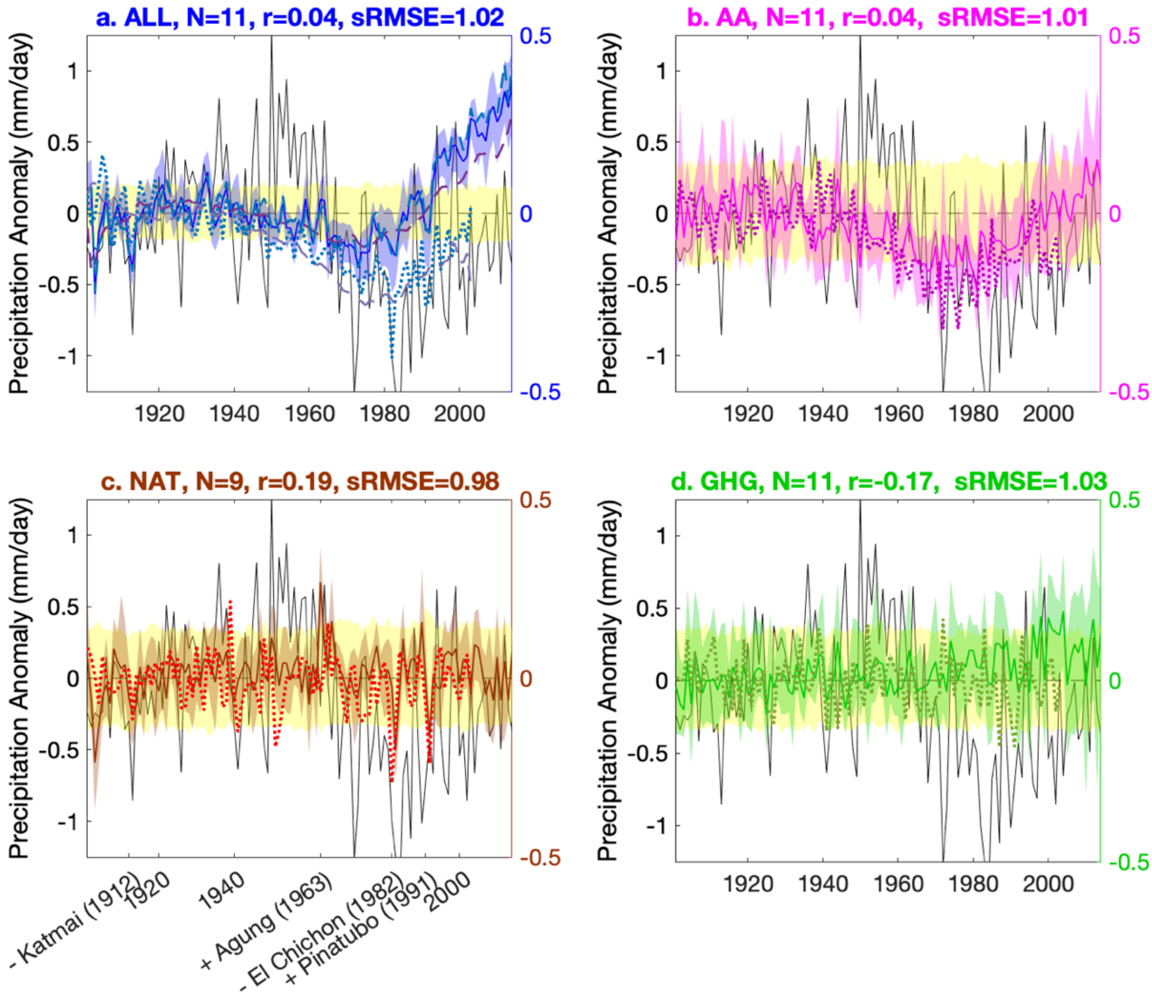

Fig. 2. Observed (black, left ordinates) and simulated (colored, right ordinates) Sahelian precipitation anomalies, forced with ALL (a, blue), AA (b, magenta), NAT (c, brown/red), and GHG (d, green). The CMIP6 MMMs are presented with solid curves surrounded by shaded areas demarking the bootstrapping confidence interval, while the CMIP5 MMMs are presented with dotted curves. The yellow shaded area is the confidence interval of randomized bootstrapped MMMs of CMIP6 piC simulations, and represents the magnitude of noise in the CMIP6 MMMs. Hemispherically asymmetric volcanic forcing from Haywood et al noted in panel (c). A negative sign denotes an eruption that cooled the northern hemisphere more than the southern hemisphere while a positive sign denotes the opposite, aligning with the sign of the expected Sahelian precipitation response to the eruption. Panel (a) additionally shows the CMIP6 ALL MMM when restricted to models, rather than institutions, that provide AA simulations (blue dashed curve), and a 20-year running mean of the sum of the AA, NAT, and GHG MMMs for CMIP5 (lavender dashed curve) and CMIP6 (burgundy dashed curve). The label shows the number of institutions used for each CMIP6 MMM (N), the correlation of the CMIP6 MMM with observations (r), and the standardized root mean squared error of the CMIP6 MMM with observations (sRMSE).

In the AA experiments (panel b), CMIP6 is anomalously wetter than CMIP5 in the 1970s and around 2000, but otherwise looks similar to CMIP5: precipitation declines in the mid-century and then recovers after the clean air acts, preceding the timing of observed variability by about 10 years. There are some differences in the NAT experiments between CMIP5 and CMIP6 (panel c), but the largest variations in both ensembles are interannual episodes that are clearly associated with volcanic eruptions. In the GHG experiments (panel d), CMIP6 shows anomalous wetting after 1970 that wasn't present in CMIP5.

Similar changes can be seen in the ALL simulations (panel a): while CMIP5 reaches peak drought in 1982 – close to the observed precipitation minimum – CMIP6 dries very little and only until 1970, after which it displays an anomalously wetter climate than CMIP5 through the end of the century. But while the precipitation responses to different forcing agents appear to add linearly in CMIP5 (compare the lavender dashed curve to the blue dotted curve), the late century wetting in CMIP6 is larger than the sum of GHG and AA wetting (burgundy dashed curve; including NAT does not help.) This effect is robust to differences in model availability for the different sets of forcing agents (see figure caption and light blue dashed curve). Thus, in the ALL simulations, CMIP6 displays slightly less drying from AA compared to CMIP5, more wetting from GHG, and additional wetting after 1990 from a non-linear interaction between forcings.

As a result of these changes, the response to forcing in CMIP6 is a poor match to observations. Figure 3 displays the correlation (panel a, "r") and sRMSE (panel b) between observations and simulated MMMs (dots) and bootstrapped MMMs (curves) from CMIP6 (ALL in blue, AA in magenta, NAT in brown, and GHG in green solid curves) and CMIP5 (ALL and AA in blue and magenta dotted-dashed curves; other simulations omitted for clarity) from 1901 to the end of the simulations (2003 for CMIP5 and 2014 for CMIP6). The dotted curves present the randomized bootstrapping distributions for the CMIP6 piC simulations, and the vertical dashed lines mark the one-sided $p=0.05$ significance level given by these distributions. Recall that correlation measures similarity in timing between simulations and observations where 1 is a perfect match, and sRMSE measures the amplitude of differences between the simulations and observations where 0 is a perfect match.

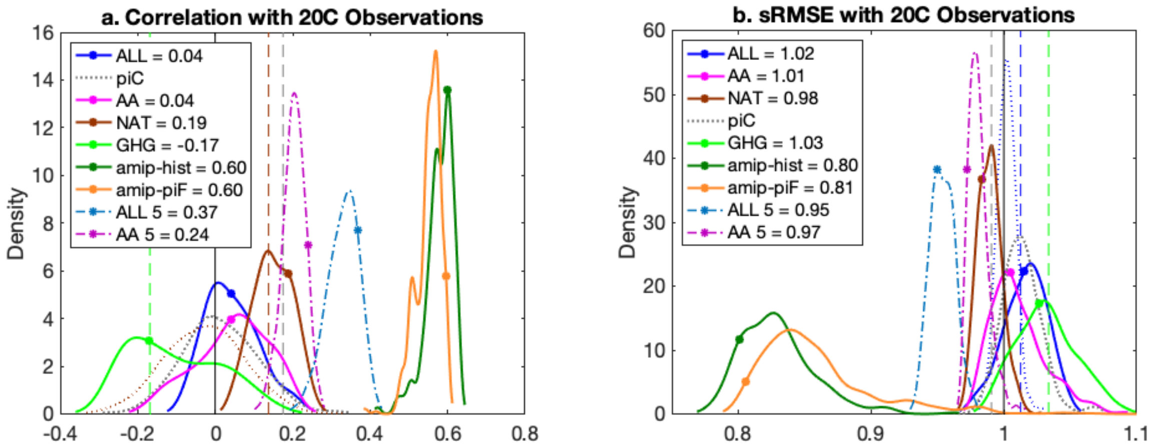

Fig. 3. Correlations (a) and standardized RMSE (b) between observations and historical and AMIP simulations from CMIP6 (1901-2014, solid) and those simulations from CMIP5 that outperform the CMIP6 historical simulations (1901-2003, dotted-dashed, legend entries include “5”). Dots and stars denote the statistic between the MMM and observations, while the curves denote the bootstrapping pdfs. The dotted grey curves display the bootstrapping pdfs for the same statistics applied to a MMM over the CMIP6 piC simulations, and the grey dashed lines mark the one-sided p=0.05 significance level given by the piC distribution. Colored dotted curves and dashed lines show the piC distributions associated with those subsets of forcing agents for which the piC distribution differs noticeably from those of the other subsets of forcing agents.

CMIP5’s AA (r = 0.24, sRMSE = 0.97) and ALL (r = 0.37, sRMSE = 0.95) MMMs achieve significance in both metrics – a fact that, in isolation, is consistent with the suggestion that AA may explain observed variability but underestimate its magnitude. Instead, in CMIP6, AA (r = 0.04, sRMSE = 1.01) and ALL (r = 0.04, sRMSE = 1.02) do not perform statistically better than noise, and GHG performs significantly worse (r = -0.17, sRMSE = 1.03). The additional years included in the CMIP6 simulations (2004-2014) cannot explain the entire deterioration of performance between CMIP5 and CMIP6: even when restricted to CMIP5’s time period, CMIP6 ALL and AA simulations both perform worse than CMIP5 in both metrics (r = 0.07 and sRMSE = 1.00 for AA, r = 0.13 and sRMSE = 0.99 for ALL). Most of the remaining deterioration in performance for AA is due to reduced drying in the 1970s in CMIP6. In CMIP6, NAT (r = 0.19, sRMSE = 0.98) is the only forcing that performs significantly well. We conclude that aside from episodic responses to volcanic eruptions, the ensemble of coupled CMIP6 simulations has no significant skill in simulating historical Sahel rainfall in response to external forcing.

As in CMIP5, the simulated forced component of precipitation changes in CMIP6—given by the MMM—has a much smaller variance than observations (note the amplification of the right ordinates in Figure 2). However, the poor performance of the CMIP6 simulations makes it clear that amplifying the simulated forced component will not help explain observed precipitation.

For simulated atmospheric and oceanic IV ($\vec{a}$ and $\vec{o}$) to explain observed precipitation variability, it is not enough that observed yearly Sahelian precipitation anomalies fall within the range of individual simulations (not shown)—the latter must also match the distinctive low-frequency power of the observations. In Figure 4 we compare the power spectra (PS) of piC simulations (colored brown to turquoise by model climatological rainfall) to the observed PS (solid black) and the PS of the ALL-residual (observations minus the ALL MMM, dotted-dashed black). In the observations and the residual, variance at periods longer than about 20 years (low-frequency) is roughly 5 times as large as the high-frequency variance. Low-frequency variability in the piC simulations is smaller than, and inconsistent with, either observed or residual variability. Moreover, it is similar in magnitude to simulated high frequency variability, suggesting that IV in simulated Sahel rainfall derives mostly from atmospheric ($\vec{a}$), rather than oceanic ($\vec{o}$), IV, or that simulated oceanic IV is too white (Eade et al. 2021). Because the shape of the spectrum is wrong, even a bias correction that inflates simulated internal variability would not bring simulations and observations into alignment.

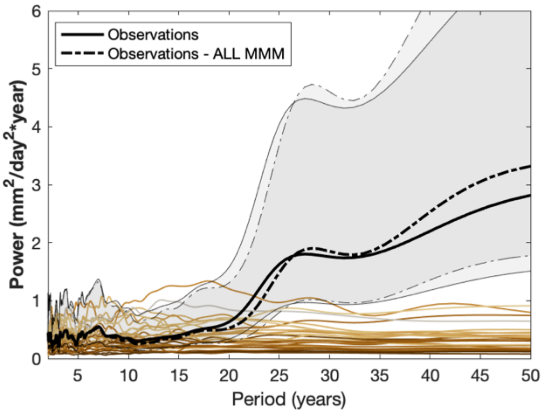

Fig. 4. PS of observed Sahelian precipitation (solid black curve) and the residual of observations and the ALL MMM (dotted-dashed black curve) and associated 95% confidence intervals (grey shading), compared to the average PS by model of piC simulations (brown to turquoise). Mean piC PS are colored by the average yearly piC precipitation by model, where brown simulations are drier than observed, and turquoise simulations are wetter than observed.

We must conclude that no linear combination of the simulated forced signal (which correlates poorly with observations) and simulated IV (which has insufficient low-frequency variance) in coupled CMIP6 simulations can explain observed Sahel variability during the 20^{th} century. Thus, model deficiency cannot be blamed solely on the simulation of climate feedbacks: the CMIP6 ensemble displays a fundamental inability to simulate the observed fast and slow Sahelian precipitation responses to forcing, observed low-frequency IV, or both. To identify the proximate cause of this failure, in the next three sections we examine each causal path component identified in Figure 1.

b. AMIP simulations: the Response to SST, Atmospheric Internal Variability, and the Fast Response to Forcing ($\vec{t}$, $\vec{a}$, and $\vec{f}$)

To isolate the effect of SST on the Sahel ($\vec{t}$), we examine precipitation in the CMIP6 amip-piForcing simulations, which force atmosphere-only models with the observed SST history (containing both internal, $\vec{o}$, and forced, $\vec{s}$, oceanic variability) and constant preindustrial external radiative forcing (no $\vec{f}$). The MMM of simulated Sahel precipitation filters out atmospheric IV ($\vec{a}$), leaving the precipitation response to the entire observed SST field. It is displayed in Figure 5a (orange) and compared to observations (black) on the same ordinates. Overall, the performance of the amip-piF MMM is much better than that of the coupled simulations: it achieves a high correlation (r = 0.60) and a low sRMSE (0.81, see orange curves in Figure 3). The good match with observations is achieved mostly at low frequencies: though it doesn't accurately capture many interannual episodes—notably including the precipitation minimum in 1984—the MMM appears to capture the magnitude of low-frequency variability, even including wetting in the 50s and early 60s, which is missing from the coupled MMM. This can be seen more quantitatively by spectral analysis. In Figure 6a, the PS of the amip-piF MMM (dashed orange curve) and its 95% confidence interval (orange shaded areas), are compared to those of observations (black). Unlike previous generations of AMIP experiments (e.g. Scaife et al. 2009), the PS of the simulated MMM is roughly consistent with observations.

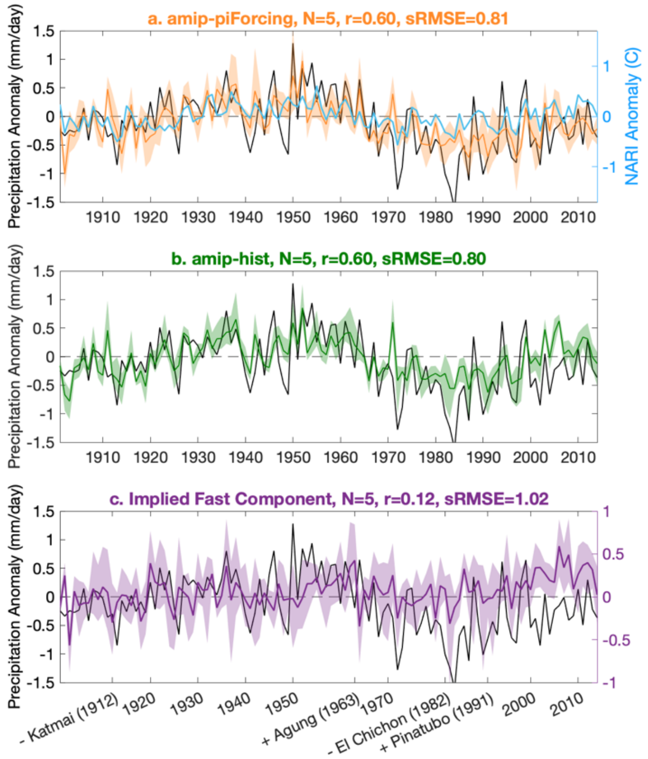

Fig. 5. Observed (black) and simulated (colored) Sahelian precipitation anomalies, forced with observed SST alone (a, amip-piF, orange) and with observed SST and all external forcing agents (b, amip-hist, dark green). The shaded areas denote the bootstrapping confidence intervals about the simulated MMMs. Panel (a) additionally displays observed NARI (light blue, right ordinates). The right ordinates for panel (a) are scaled by the inverse of the simulated amip-piF teleconnection strength (see Section 4.c) so that when read on the left ordinates, NARI represents its predicted impact on precipitation. Panel (c) compares observed precipitation (left ordinates) to the implied simulated fast component in AMIP simulations (amip-hist – amip-piF, purple, right ordinates). As in Figure 2, panel (c) denotes hemispherically asymmetric volcanic eruptions, where the sign denotes the sign of the expected Sahelian precipitation response to the eruption.

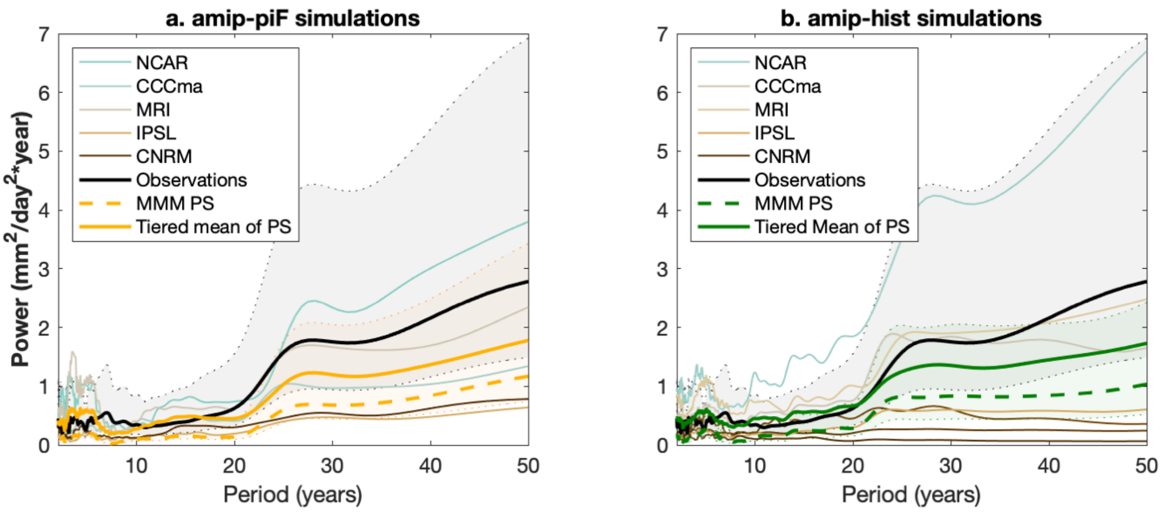

Fig. 6. PS of observed Sahelian precipitation (black) and associated 95% confidence interval (black shading) compared to the PS of amip-piF simulations (a) and amip-hist simulations (b). As in Figure 4, mean PS by model are colored by average yearly precipitation, where brown is drier than observed, grey is observed, and turquoise is wetter than observed. The mean PS across models is displayed in orange for amip-piF (a) and in green for amip-hist (b). The dashed lines show the PS of the MMMs with associated 95% confidence intervals (colored shaded areas).

The curves colored brown to turquoise in Figure 6 show the average by model of the PS of individual simulations, colored by climatological Sahelian precipitation bias. We note that wet-biased simulations (turquoise) have more power than dry-biased simulations (brown), consistent with the expected relation between the mean and variance of precipitation. The tiered mean over these PS is presented in solid orange; it contains atmospheric IV ($\vec{a}$) in addition to SST-forced variability ($\vec{t}$). Though it is not statistically different from the MMM PS, atmospheric white noise gives it slightly more power at all frequencies, and thus it is clearly consistent with the observed PS (black). Global SST forcing, while unable to explain much of observed high frequency variability in Sahelian precipitation (note the low power of the dashed orange curve at periods below 20 years), is able to reproduce the pattern and, in combination with atmospheric IV, the full magnitude of observed multi-decadal precipitation variability.

We now estimate the "fast" precipitation response to ALL in the CMIP6 AMIP simulations (Figure 5c, purple, $\vec{f}$) by subtracting the MMM of amip-piF simulations (a, orange) from that of amip-hist simulations (b, green), the latter of which are forced with historical SST and historical external radiative forcing. The AMIP "fast" MMM shows some episodic variability that is consistent with the coupled NAT MMM, and a wetting trend after 1985. On its own, it is only weakly correlated to observations ($r = 0.12$, sRMSE = 1.02), and it has relatively low amplitude. When combined with SST forcing in the amip-hist simulations, it has little effect: correlation stays at 0.60 and sRMSE is reduced from 0.81 only to 0.80 (compare green and orange curves in Figure 3) and spectral properties are virtually unchanged (Figure 6). The best linear fit to observed precipitation would combine the amip-piF MMM with the fast response to forcing scaled down by a factor of 0.3 ± 0.2. The fast response may be overestimated in AMIP simulations because the radiative forcing has directly contributed to generating observed SST which is prescribed in the simulations, and because the magnitude of the radiative forcing itself may be overestimated, as suggested by Menary et al. (2020).

The high performance of the amip-piF simulations and the small impact of the potentially overestimated fast response to forcing suggest that the principal deficiency in simulating low-frequency Sahelian precipitation variability in coupled models stems from a deficiency in simulating the observed combination of forced and internal variability in SST, and not from a failure to reproduce the observed teleconnection strength or fast response to forcing.

c. The NARI Teleconnection: AMIP Simulations and Observations ($\vec{t}$)

We next determine the strength of the linear NARI-Sahel teleconnection and investigate how well it represents the effect of global SST on Sahel precipitation in simulations and observations. Observed NARI anomalies relative to the 1901-1950 mean are presented in Figure 5a in light blue on the right ordinates. NARI correlates well with SST-forced Sahelian precipitation in the amip-piF simulations (orange, left ordinates; $\mathrm{r} = 0.52 \pm 0.10, \mathrm{r} = 0.60$ for the actual MMM), but still leaves 64% of its variance unexplained, suggesting influences from other SST patterns or non-linear or non-stationary effects (Losada et al. 2012). Some of the unexplained variance is at faster timescales than those of our interest, but not all. Let's assume that the influences of NARI and other ocean basins on Sahel precipitation are linear and add linearly, and that the NARI teleconnection is unconfounded by the influence of other ocean basins; then we can measure the strength of the NARI teleconnection by the regression coefficient of the amip-piF precipitation MMM, which contains only SST-forced variability, on NARI. This calculation yields a regression slope of $0.87 \pm 0.26 \frac{\mathrm{mm}}{\mathrm{day} * {}^{\circ}\mathrm{C}}$. This value is affected by both high- and low-frequency variability, which is appropriate if the teleconnection is, indeed, linear. The left ordinates in Figure 5a are scaled relative to the right ordinates by this teleconnection strength so that, when read on the left ordinates, the light blue curve represents the expected precipitation response to NARI. This view highlights how NARI captures the timing of simulated low-frequency variability, even though it fails to explain the full magnitude of simulated dry anomalies after 1975. In the rest of this paper we use the NARI teleconnection as the best linear representative of the simulated influence of SST on Sahel precipitation in the 20th century.

The teleconnection strength calculated from the amip-piF simulations is not directly comparable to observations, because the latter includes the fast precipitation response to forcing, which can confound estimates of the teleconnection. A comparison can be drawn between the apparent teleconnection strength in the amip-hist simulations (0.93 ± 0.41) and in observations (1.04). The consistency lends credence to our previous suggestion that simulated SST teleconnections to Sahel rainfall appear to have the appropriate strength in CMIP6, at least in the amip simulations.

d. Forced and Internal SST Variability in Coupled Simulations ($\vec{s}$ and $\vec{o}$)

We now examine simulation of forced ($\vec{s}$) and internal ($\vec{o}$) SST variability. Figure 7 compares observations (black) to the simulated SST response to forcing ($\vec{s}$)—represented by MMM anomalies (colors)—for NARI (right column) and its constituent ocean basins – the North Atlantic (NA, left column) and the Global Tropics (GT, middle column). The yellow shaded areas show the bootstrapping 95% confidence intervals of the piC simulations for statistical significance, while the other shaded areas denote uncertainty in the CMIP5 and CMIP6 MMMs. As above, CMIP5 MMM anomalies are presented in dotted curves and CMIP6 in solid curves, color-coded according to their forcing.

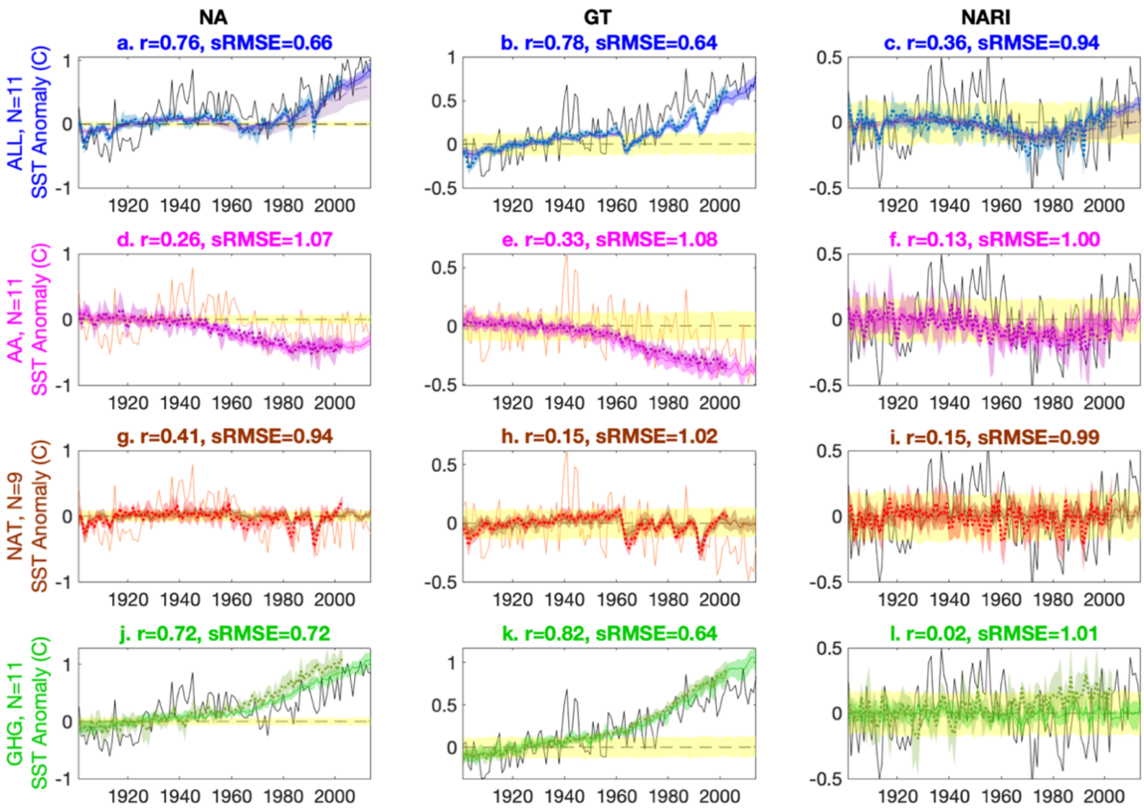

Fig. 7. Observed (black) and simulated CMIP5 and CMIP6 SST anomalies (relative to 1901-1950) for the North Atlantic (NA, left column), the Global Tropics (GT, middle column), and the North Atlantic Relative Index (NARI, right column) when forced with ALL (blue, top row), AA (magenta, second row), NAT (brown/red, third row), and GHG (green, bottom row). The CMIP6 MMMs are presented with solid curves while the CMIP5 MMMs are presented with dotted curves. Both are surrounded by shaded areas demarking the

bootstrapping confidence interval. Panels (a) and (c) additionally display a 20-year running mean of the sum of simulated NA and NARI over the individual forcing simulations for CMIP6 (burgundy dashed curve) with associated bootstrapping confidence interval (burgundy shaded area). Including NA in the sum makes little difference. For NA and GT under AA and NAT (middle two rows and left two columns), the orange curve displays detrended observations, calculated by subtracting simulated GHG-forced SST (bottom row) from observations in that ocean basin. The yellow shaded area is the confidence interval when bootstrapping the MMM of CMIP6 piC simulations, and represents the magnitude of noise in the CMIP6 MMMs. A horizontal black dashed line marks 0 anomaly, which represents the average SST from 1901-1950. The y labels show the number of institutions that were used for each subset of forcing agents in CMIP6 (N, see Table S2), and the subplot titles display the correlation (r) and sRMSE between the MMM and observations for CMIP6.

Observed NARI (panel c, black) shows strong multi-decadal variability throughout the century. In the ALL simulations (top row, blue), the temporal evolution of NARI (c) matches the observations with some skill (r=0.40, sRMSE = 0.92 for CMIP6), but fails to capture the full magnitude of observed cooling in the 1970s and 80s or, more prominently, any multi-decadal variability prior to 1960. Moreover, its GT and NA components do not match very well either the observed, roughly linear warming trend in GT (b), or the marked multi-decadal variability in NA (a). In both CMIP5 and CMIP6 ALL simulations, the simulations of GT (b, blue) are anomalously colder than observations between 1960 and 2000, when simulated AA cooling (e, magenta) is the strongest and not yet compensated by GHG warming (k, green), leading us to question whether the match of simulated and observed NARI in this period happens due to compensating errors. For NA, the match between observations and the ALL-forced response is better in the later part of the record, but worse in the first half. During the period prior to 1960, according to both CMIP ensembles, GHG warming (j, green) masks AA cooling (d, magenta) to produce a roughly constant temperature in the ALL simulations (a, blue). The simulated cold episode in 1964 is due to the eruption of Agung in 1963 (g, brown and red), and it is only after the mid 1960's that increased GHG warming overtakes stagnating AA cooling to produce pronounced warming in fairly good accord with observations. Much of the observed variability in NA (a, black) thus does not seem to be a response to external radiative forcing.

The AA forcing had appeared to explain observed low-frequency Sahel precipitation variability in H20, but we now see that it might be the right result for the wrong reason. AA (second row, magenta) produce low-frequency NARI variability (f), but this simulated NARI is a poor match to observations (f, r=0.10, sRMSE = 1.04 for CMIP5; r=0.07, sRMSE=1.09 for CMIP6; a performance statistically worse than noise). The difference between simulations and observations is even more stark in NARI's constituent ocean basins. We can attempt to compare AA-forced NA and GT to an observed "GHG-residual" (that is, the observation minus the GHG-forced MMM, presented in orange instead of black), which represents our best estimate of the sum of observed oceanic IV and the observed responses to aerosols. This index shows marked, roughly stationary low-frequency variability in NA (d, orange), which contrasts with a more monotonic behavior in the simulated NA index (magenta). In particular, we note that the AA simulations display an especially steep decline in NA SST between ~1940 and 1980, but monotonic cooling throughout the century. Though legislation to curb pollution reduced AA loading in the northern hemisphere after 1970 (Hirasawa et al. 2020), simulated NA doesn't warm at all before 2010. Overall, the effect of reducing AA emissions in both CMIP ensembles is to halt the cooling of NA, not to cause actual warming. This is consistent with estimates of the hemispheric difference in total absorbed solar radiation in AA simulations in CMIP6, which level off, but do not decrease, after 1970 (Menary et al. 2020).

Could internal SST variability ($\vec{o}$) explain the difference between the simulated response to forcing and observations in these ocean basins? In Figure 8, we present the mean PS of SST for piC simulations from each CMIP6 model (colder than observed models are in blue and warmer than observed models are in red). We compare these PS to the PS for observed SST (solid black), the GHG-residual (dotted-dashed black), and/or the ALL-residual (dotted black), avoiding time series with dramatic trends (see subplot legends). Simulated IV in most of the CMIP6 models used in this study does not match residual or observed low-frequency variability in NA (a), GT (b), or NARI (c). In CMIP5, SSTs are colder and IV at all frequencies is larger than in CMIP6, but no model shows an increase in spectral power at low frequencies for any SST index (not shown). There are, however, three CMIP6 models for which low-frequency IV in NA is not inconsistent with model physics: CNRM-ESM2-1 p1 (pink), IPSL-CM6A-LR p1 (blue), and CNRM-CM6-1 p1 (grey). Certainly, either the simulated SST response to forcing, simulated oceanic internal variability, or both, are not well represented in the CMIP ensembles, and this is the primary reason that coupled CMIP simulations cannot reproduce observed 20th century Sahel rainfall.

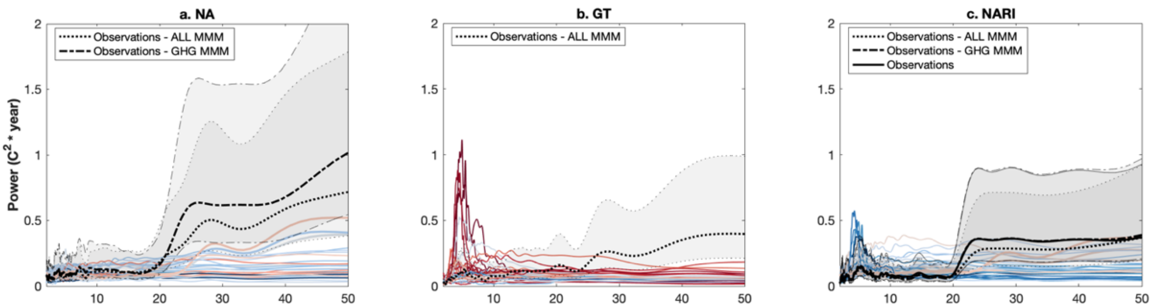

Fig. 8. PS of observed SST (solid black), observed SST – GHG MMM (dotted-dashed black), observed SST – ALL MMM (dotted black) and associated 95% confidence intervals (black shading) in NA (a), GT (b), and NARI (c), compared to the PS of piC simulations. Similar to Figure 4, mean PS by model are colored by average SST, where blue is colder than observed, grey is observed, and red is warmer than observed.

However, deficiencies in simulating SST cannot explain the difference in simulated externally forced precipitation variability between CMIP5 and CMIP6. The only notable difference in simulated SST between the two ensembles is that CMIP6 warms NA (and therefore NARI) less than CMIP5 in the GHG simulations (Figure 7j and l). As in simulated Sahel precipitation, warming of NA and NARI in CMIP6 ALL simulations is larger than the smoothed sum of simulated SST change in the individual-forcing simulations (burgundy dashed curve), which, aside from volcanic eruptions, remains below the confidence interval for the CMIP6 MMM (dark blue shaded area) from 1950 onward (this discrepancy is, again, robust to differences in model availability for the different sets of forcing agents). Thus, a non-linear interaction between forcing agents in CMIP6 balances the additional SST warming in CMIP5 in the ALL simulations, and the difference in coupled simulations of Sahel rainfall between CMIP5 and CMIP6 must derive from changes in the fast response to forcing, SST teleconnections, or both.

e. The NARI teleconnection in Coupled Simulations

Now that we have examined SST in the coupled simulations, we may determine whether the teleconnection strength estimated from amip-piF simulations is consistent with coupled simulations. This is verified by the fact that the amip-piF teleconnection strength falls within the range of teleconnection strengths calculated from individual piC simulations in CMIP5 (0.5 ± 0.6) and CMIP6 (0.4 ± 0.6), but the ranges are large (possibly because the increased presence of atmospheric and oceanic IV and decreased variance of NARI in the individual piC simulations obscures the teleconnection). As a second test, we compare the confounded teleconnection strength in the amip-hist simulations (0.93 ± 0.41) to that of bootstrapped

MMMs in the coupled ALL simulations in CMIP5 (0.66 ± 0.28) and CMIP6 (1.5 ± 0.3). The confounded teleconnection strength in amip-hist simulations is consistent with the confounded estimate in CMIP5, but is smaller than and inconsistent with the confounded estimate in CMIP6. This may be because NARI variability in the coupled simulations is smaller relative to the magnitude of external radiative forcing than it is in the amip-hist simulations. If this is the cause for the apparent inconsistency, we may still confirm the NARI teleconnection strength in CMIP6 simulations by showing that the implied fast response to forcing is consistent with the fast response from the amip-hist simulations.

f. Fast and Slow Responses to Forcing in Coupled Simulations ($\vec{f}$ and $F \rightarrow SST \rightarrow P$)

Under the assumption that the dominant simulated path of SST influence on the Sahel is captured by a linear relationship with NARI, we estimate the slow response to forcing in coupled simulations as the simulated NARI MMM scaled by the teleconnection strength derived from uncoupled simulations ($0.87 \frac{\mathrm{mm}}{\mathrm{day}\ {}^{\circ}\mathrm{C}}$, Section 4.c), so that a warm (cold) NARI predicts a wet (dry) Sahel. In Figure 9, simulated NARI (as in Figure 7, right column) is displayed on the left ordinates in light blue (CMIP6) and turquoise (CMIP5). The right ordinates are scaled by the teleconnection strength so that, when read on the right ordinates, simulated NARI represents the estimated slow component of the precipitation response to forcing. Also on the right ordinates are the total simulated precipitation responses to forcing (as in Figure 2) in CMIP5 (right column) and CMIP6 simulations (left column), colored by forcing agents. The simulated precipitation responses to forcing (colors) match the estimated slow response to forcing (turquoise) reasonably well: the main differences appear after about 1970 in CMIP5 and 1990 in CMIP6.

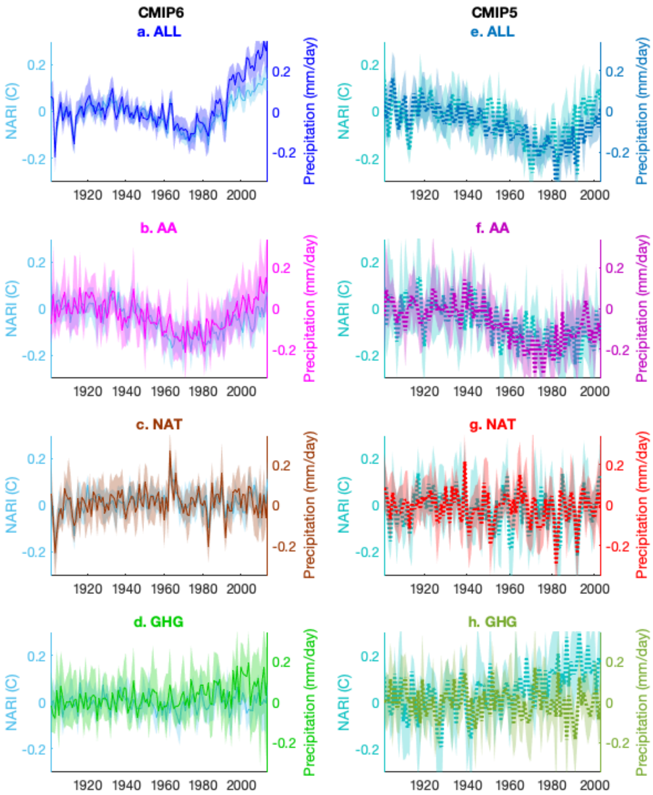

Fig. 9. Simulated Sahel precipitation (right ordinates, same as Figure 2) MMMs (solid and dotted curves) and associated 95% confidence intervals (shaded areas) in CMIP5 (right column) and CMIP6 (left column) when forced with ALL (blue, top row), AA (magenta, second row), NAT (brown/red, third row), and GHG (green, bottom row), compared to simulated NARI (left ordinates, light blue and turquoise, same as Figure 7). The right ordinates are scaled such that a 1°C change in NARI corresponds to a 0.87 mm/day change in precipitation, given by the teleconnection strength in the CMIP6 amip-piF simulations (see Section 4.c).

We expect the differences between the simulated Sahel and the rescaled NARI to estimate the simulated fast response to forcing, but this would imply a fast response to ALL in CMIP5 (Figure 9e) that is inconsistent with the uncoupled estimate (purple, Figure 5c): instead of wetting the Sahel, it consists of a drying response to increasing GHG of $-0.0042 \pm 0.0036 \frac{\mathrm{mm}}{\mathrm{day}*\mathrm{year}}$ (Figure 9h). Whether we should interpret this as a fast response or a non-NARI-mediated response to SST, this component of the forced response helps delay and increase the severity of the minimum in precipitation in ALL relative to the AA simulations.

The estimated fast responses for CMIP6 are displayed in Figure 10 in a fashion similar to Figure 2, and are compared to the fast response obtained as the difference between amip-hist and amip-piF simulations (purple, as in Figure 5c). Unlike the fast response in CMIP5, the ALL fast response in CMIP6 matches the AMIP fast response significantly better than noise (r = 0.51, sRMSE = 0.87), giving us confidence that the NARI teleconnection strength estimated from amip-piF is valid in CMIP6 coupled simulations. Like the amip-hist fast response, the ALL fast response in CMIP6 displays wetting after 1980 that is roughly equal to the sum (burgundy dashed curve) of the fast responses to AA (b, magenta) and GHG (d, green). The simulated fast wetting after 1980 in the ALL simulations (a, blue) is smaller than in the AMIP simulations, as expected if amip-hist is double-counting radiative forcing, but is still larger than our estimate of the optimal value (0.3 times the AMIP fast response), consistent with claims that the strength of radiative forcing is overestimated in the coupled simulations.

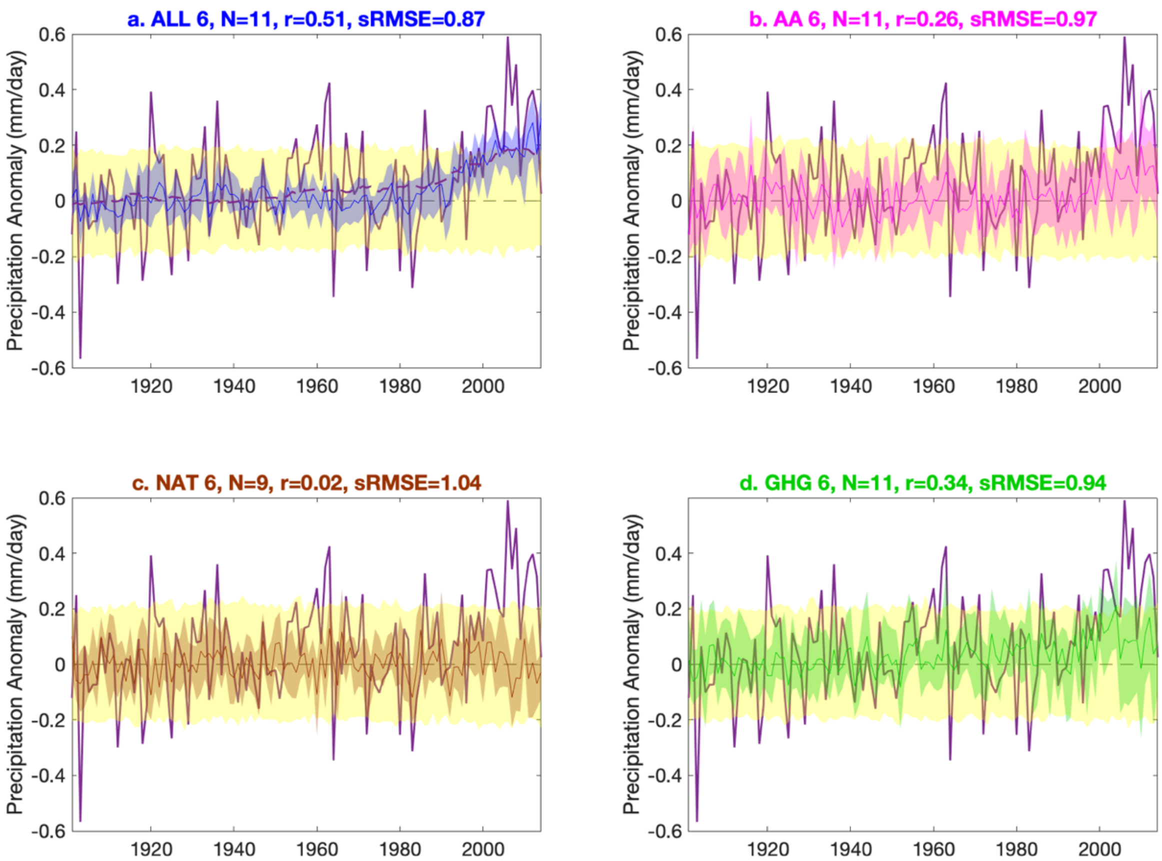

Fig. 10. Compares the fast Sahelian precipitation response to forcing in AMIP simulations (purple, as in Figure 5c) to the estimated fast component of the precipitation MMMs in coupled CMIP6 simulations (precipitation – 0.87*NARI; the difference between the colored and light blue curves in the left column of Figure 9) forced with ALL (a, blue), AA (b, magenta), NAT (c, brown), and GHG (d, green). Similar to Figure 2, the colored shaded areas denote the bootstrapping confidence interval of this difference, and the yellow shaded areas, which represent the magnitude of noise in the fast MMMs, are the confidence intervals of the MMM of randomized bootstrapped differences between precipitation and 0.87*NARI in piC simulations. Panel (a) additionally shows a 20-year running mean of the sum of the AA, NAT, and GHG fast MMMs (burgundy dashed curve). The label shows the number of

institutions used for each CMIP6 MMM (N), the correlation of the fast MMM with the AMIP fast response (r), and the standardized root mean squared error of the CMIP6 MMM with observations (sRMSE).

Though NARI in the GHG simulations differs between CMIP5 and CMIP6, most of the difference in simulated forced precipitation between CMIP5 and CMIP6 is not mediated by a linear relationship with NARI, and can be attributed to the fact that the GHG- and AA-induced drying in CMIP5 is replaced with AA- and GHG-induced wetting in CMIP6. Whether the GHG-induced drying in CMIP5 is a fast response to forcing or a response mediated by SST in ocean basins other than the Atlantic cannot be firmly established by this analysis, but we offer our perspective below.

5. Discussion

Using SST (and specifically NARI) as a mediator, we have established that the failure of CMIP coupled models to simulate observed Sahel rainfall stems from their inability to simulate observed SST, especially NA, and that the differences in simulation of Sahel rainfall between CMIP5 and CMIP6 stem from differences in mechanisms not mediated by a linear teleconnection with NARI. (Let's denote the difference between simulated precipitation and scaled NARI as $P_{nonNARI}$). We initially suggested that $P_{nonNARI}$ provides a good measure of the fast (non-SST-related) response to forcing because of the prominence of the NARI-Sahel teleconnection in observations and AMIP-style simulations of the 20th century. But without examining further mediators, we cannot decisively rule out the possibility that $P_{nonNARI}$ captures teleconnections with other ocean basins or nonlinearities in the NARI teleconnection. Which explanation is most likely?

The $P_{nonNARI}$ indices in CMIP5 and CMIP6 are nearly opposite. If we assume that both represent a fast response to forcing, we need to conclude that increasing GHG (or reducing AA) lead to fast wetting in CMIP6, but drying in CMIP5.

The interpretation of $P_{nonNARI}$ in CMIP6 as a fast response is more consistent with theory. First, increasing rainfall is consistent with theory linking reduced aerosol concentrations to fast surface warming and decreasing optical depth of the atmosphere (Allen and Ingram 2002; Rosenfeld et al. 2008), although a couple highly non-linear simulations suggest the fast precipitation response of the Sahel to changing AA in the 20th century was drying whether AA forcing was increasing or decreasing (Hirasawa et al. 2020). Second, it is generally accepted that the fast response of the Sahel to GHG is wetting (e.g. Biasutti 2013; Gaetani et al. 2017; Giannini 2010; Haarsma et al. 2005). The good match in the estimated fast response between coupled CMIP6 simulations and the amip-hist simulations increases our confidence that the deviations from the NARI-mediated slow response to forcing in CMIP6 really reflect a fast response to forcing. The same cannot be said for CMIP5.

We noted in Section 4.c that NARI only explains 36% of simulated SST-forced variability in the amip-piF simulations, leaving room for the influence of other ocean basins or SST indices on Sahel precipitation. Indeed, this is consistent with GK19: while they argue that NARI is the primary indicator for 20th century Sahel rainfall, they also argue that p1, which is approximately (NA+GT)/2 and is intended to capture the effects of uniform global warming, plays a secondary—but important—role in the 20th century and a dominant role in the future. In CMIP5, $P_{nonNARI}$ may capture not the fast responses to forcing, but slow drying in response to uniform global warming, consistent with previous literature (e.g. Gaetani et al. 2017). In this read, the differences in simulation of Sahel rainfall between CMIP5 and CMIP6 are due to a combination of changes in the fast response to forcing and the influence of SST patterns not captured by NARI.

6. Summary and Conclusions

In this paper, we decompose simulated Sahelian precipitation into (1) teleconnections with SST, (2) fast, atmospheric- and land-mediated responses to forcing, (3) atmospheric noise, (4) forced SST variability, and (5) internal SST variability, in order to determine why the 5th and 6th generations of CMIP differ in their simulation of Sahel rainfall, and why both ensembles are inconsistent with observed Sahel precipitation variability.

CMIP6 atmospheric simulations forced with observed SST alone capture observed Sahel precipitation quite well (r=0.6), and, in combination with atmospheric white noise, are able to reproduce the power of observed low-frequency variability. This is a welcome improvement from previous generations of climate models. Including radiative forcing alongside observed SST barely changes simulated precipitation, suggesting that the fast response is small and plays a secondary role to SST-forced precipitation variability. We summarize the Sahel teleconnections with global SST as a linear relationship with an index of the warming of the North Atlantic relative to the global Tropics (NARI), which explains about 36% of the simulated precipitation response to observed SST. The simulated NARI teleconnection is measured as $0.87 \pm 0.26 \frac{\mathrm{mm}}{\mathrm{day} * {}^{\circ}\mathrm{C}}$, consistent with the strength of the observed teleconnection.

We conclude that the observed SST history and simulated teleconnections in atmospheric simulations are together necessary and sufficient to capture the timing and magnitude of the low-frequency droughts and pluvials in 20th century Sahel rainfall.

In coupled simulations, the NARI-Sahel teleconnection is consistent with AMIP simulations, but NARI's variability – which mostly comes from North Atlantic SST (NA) – differs from the observed. In simulations, AA cause a cooling trend and GHG cause a warming trend with magnitudes comprable to the observed, but no combination of forcing agents produces a decadal-scale oscillation in NA in either CMIP5 or CMIP6, and only three CMIP6 models (out of 25 CMIP5 and 30 CMIP6 models) are able to generate internal SST variability commensurate to the residual (the difference between total and radiatively forced) low-frequency variability. How do we reconcile our results with those claiming that the observed Atlantic Multidecadal Variability (AMV) is externally forced (mainly by AA; Bellomo et al. 2018; Booth et al. 2012; Hirasawa et al. 2020; Hua et al. 2019; Murphy et al. 2017)? The discrepancy can be explained because these studies examine only one or two models (Booth et al. 2012; Hirasawa et al. 2020) or subtract a linear trend from simulated NA before comparing to observations (Bellomo et al. 2018; Hua et al. 2019; Murphy et al. 2017), thus inducing low-frequency variability in the simulated monotonic decreasing step function. Moreover, a prominent role for internal variability cannot yet be dismissed, as suggested by Yan et al. (2018), who, consistent with our analysis, find that most models do not capture observed AMOC variability. The NARI-mediated slow response to external radiative forcing is to dry the Sahel slightly in the 60s and to wet it immediately afterwards; this does not, in isolation, explain the timing or magnitude of the observed drought or recovery. Furthermore, forced NARI variability is small in the first half of the century. We are led to conclude that either the pattern of the simulated SST response to forcing in coupled models is incorrect or the Sahelian precipitation response to internal SST variability overshadowed the response to external radiative forcing in the 20th century, at least up to the mid-1960s.

While we can ascribe the deficiency of 20th century Sahel rainfall simulations in both CMIP5 and CMIP6 coupled models to their simulations of SST, NARI is not the main explanation for the differences in forced Sahel rainfall between the two ensembles, since it is quite similar in CMIP5 and CMIP6 ALL simulations. The difference, rather, is in $P_{nonNARI}$: the component of Sahel rainfall that comes either from the influence of other SST patterns or from the fast response to forcing. CMIP6 underperforms relative to CMIP5 because $P_{nonNARI}$

includes substantial fast wetting responses to increasing GHG and decreasing AA, comparable in magnitude to the NARI-related component. In contrast, $P_{nonNARI}$ in CMIP5 is drying, likely in response to uniform SST warming. Sahel drying in response to uniform warming is strong in models that simulate a deeper ascent profile, but weak otherwise (Hill et al 2017), so it is possible that newer parameterizations and higher resolution have changed the sensitivity to this forcing in the latest generation of models.

This work has shown that, while there has been progress in the simulation of the Sahel's response to global SST, much remains uncertain in the simulation of the pathways of Sahel multi-decadal variability, especially in the amplitude and timing of forced and natural SST anomalies in the Atlantic and in the fast and slow response of rainfall to GHG forcing. Differing mechanisms can lead to similar time evolutions in observations and simulations; to avoid this pitfall, future work should focus on evaluating in more detail the hypothesized pathways of the Sahel response to anthropogenic emissions and oceanic internal variability in order to further categorize model performance and improve predictions of the future.

Acknowledgments.

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 for producing and making available their model output (listed in Tables S1-S3 of this paper). 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. The CMIP6 data used in this study are available in Google cloud storage (https://console.cloud.google.com/storage/browser/cmip6) as a result of a grant to the Pangeo project (https://pangeo.io/). We thank Haibo Liu for preparing the CMIP5 data for use, and Naomi Henderson for transferring needed CMIP6 simulations to the cloud, and for aiding in data access and general technical support throughout the project. We additionally thank Alessandra Giannini for her guidance throughout the project. This research was supported by the U.S. National Science Foundation Grant AGS-1612904.

Data Availability Statement.

Observational data from the Global Precipitation Climatology Center (GPCC, Becker et al. 2013) and the National Oceanic and Atmospheric Administration's (NOAA) Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5, Huang et al. 2017) are freely available online (see https://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html and https://www.ncei.noaa.gov/products/extended-reconstructed-sst, respectively). CMIP5 (CMIP5, Taylor et al. 2012) and CMIP6 (Eyring et al. 2016) model data is freely available through the Earth System Grid (see https://esgf-node.llnl.gov/projects/esgf-llnl/).

REFERENCES

Ackerley, D., B. B. B. Booth, S. H. E. Knight, E. J. Highwood, D. J. Frame, M. R. Allen, and D. P. Rowell, 2011: Sensitivity of twentieth-century Sahel rainfall to sulfate aerosol and CO2 forcing. *J. Climate*, **24,** 4999-5014, https://doi.org/10.1175/JCLI-D-11-00019.1.

Adam, O., T. Bischoff, and T. Schneider, 2016: Seasonal and interannual variations of the energy flux equator and ITCZ. Part II: Zonally varying shifts of the ITCZ. *J. Climate*, **29,** 7281-7293, https://doi.org/10.1175/JCLI-D-15-0710.1.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419,** 228-232, https://doi.org/10.1038/nature01092.

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

Bellomo, K., L. N. Murphy, M. A. Cane, A. C. Clement, and L. M. Polvani, 2018: Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dynam.*, **50,** 3687-3698, https://doi.org/10.1007/s00382-017-3834-3.

Biasutti, M., 2013: Forced Sahel rainfall trends in the CMIP5 archive. *J. Geophys. Res.-Atmos.*, **118,** 1613-1623, https://doi.org/10.1002/jgrd.50206.

——, 2019: Rainfall trends in the African Sahel: Characteristics, processes, and causes. *Wiley Interdisciplinary Reviews: Climate Change*, **10**, https://doi.org/10.1002/wcc.591.

Biasutti, M., and A. Giannini, 2006: Robust Sahel drying in response to late 20th century forcings. *Geophys. Res. Lett.*, **33**, https://doi.org/10.1029/2006GL026067.

Biasutti, M., I. M. Held, A. H. Sobel, and A. Giannini, 2008: SST forcings and Sahel rainfall variability in simulations of the twentieth and twenty-first centuries. *J. Climate*, **21,** 3471-3486, https://doi.org/10.1175/2007JCLI1896.1.

Biasutti, M., and Coauthors, 2018: Global energetics and local physics as drivers of past, present and future monsoons. *Nature Geoscience*, **11,** 392-400, https://doi.org/10.1038/s41561-018-0137-1.

Bonfils, C. J., B. D. Santer, J. C. Fyfe, K. Marvel, T. J. Phillips, and S. R. Zimmerman, 2020: Human influence on joint changes in temperature, rainfall and continental aridity. *Nat. Clim. Change*, **10,** 726-731, https://doi.org/10.1038/s41558-020-0821-1.

Booth, B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484,** 228-232, https://doi.org/10.1038/nature10946.

Camberlin, P., S. Janicot, and I. Poccard, 2001: Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **21,** 973-1005, https://doi.org/10.1002/joc.673.

Caminade, C., and L. Terray, 2010: Twentieth century Sahel rainfall variability as simulated by the ARPEGE AGCM, and future changes. *Climate Dynam.*, **35,** 75-94, https://doi.org/10.1007/s00382-009-0545-4.

Chang, C.-Y., J. Chiang, M. Wehner, A. Friedman, and R. Ruedy, 2011: Sulfate aerosol control of tropical Atlantic climate over the twentieth century. *J. Climate*, **24,** 2540-2555, https://doi.org/10.1175/2010JCLI4065.1.

Dong, B., and R. Sutton, 2015: Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall. *Nat. Clim. Change*, **5,** 757-761, https://doi.org/10.1038/nclimate2664.

Donohoe, A., J. Marshall, D. Ferreira, and D. Mcgee, 2013: The relationship between ITCZ location and cross-equatorial atmospheric heat transport: From the seasonal cycle to the Last Glacial Maximum. *J. Climate*, **26,** 3597-3618, https://doi.org/10.1175/JCLI-D-12-00467.1.

Eade, R., D. Stephenson, A. Scaife, and D. Smith, 2021: Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic Oscillation? *Climate Dynam.***,** 1-14, https://doi.org/10.1007/s00382-021-05978-4.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9,** 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901–85. *Nature*, **320,** 602-607, https://doi.org/10.1038/320602a0.

Gaetani, M., and Coauthors, 2017: West African monsoon dynamics and precipitation: the competition between global SST warming and CO 2 increase in CMIP5 idealized simulations. *Climate Dynam.*, **48,** 1353-1373, https://doi.org/10.1007/s00382-016-3146-z.

Giannini, A., 2010: Mechanisms of climate change in the semiarid African Sahel: The local view. *J. Climate*, **23,** 743-756, https://doi.org/10.1175/2009JCLI3123.1.

Giannini, A., and A. Kaplan, 2019: The role of aerosols and greenhouse gases in Sahel drought and recovery. *Climatic Change*, **152,** 449-466, https://doi.org/10.1007/s10584-018-2341-9.

Giannini, A., R. Saravanan, and P. Chang, 2003: Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales. *Science*, **302,** 1027-1030, https://doi.org/10.1126/science.1089357.

Giannini, A., M. Biasutti, I. M. Held, and A. H. Sobel, 2008: A global perspective on African climate. *Climatic Change*, **90,** 359-383, https://doi.org/10.1007/s10584-008-9396-y.

Giannini, A., S. Salack, T. Lodoun, A. Ali, A. Gaye, and O. Ndiaye, 2013: A unifying view of climate change in the Sahel linking intra-seasonal, interannual and longer time scales. *Environ. Res. Lett.*, **8,** 024010, https://doi.org/10.1088/1748-9326/8/2/024010.

Haarsma, R. J., F. M. Selten, S. L. Weber, and M. Kliphuis, 2005: Sahel rainfall variability and response to greenhouse warming. *Geophys. Res. Lett.*, **32**, https://doi.org/10.1029/2005GL023232.

Han, Z., F. Luo, S. Li, Y. Gao, T. Furevik, and L. Svendsen, 2016: Simulation by CMIP5 models of the Atlantic multidecadal oscillation and its climate impacts. *Advances in Atmospheric Sciences*, **33,** 1329-1342, https://doi.org/10.1007/s00376-016-5270-4.

Haywood, J. M., A. Jones, N. Bellouin, and D. Stephenson, 2013: Asymmetric forcing from stratospheric aerosols impacts Sahelian rainfall. *Nat. Clim. Change*, **3,** 660-665, https://doi.org/10.1038/nclimate1857.

Held, I. M., T. L. Delworth, J. Lu, K. u. Findell, and T. Knutson, 2005: Simulation of Sahel drought in the 20th and 21st centuries. *Proc. Nat. Acad. Sci.*, **102,** 17891-17896, https://doi.org/10.1073/pnas.0509057102.

Herman, R. J., A. Giannini, M. Biasutti, and Y. Kushnir, 2020: The effects of anthropogenic and volcanic aerosols and greenhouse gases on twentieth century Sahel precipitation. *Scientific reports*, **10,** 1-11, https://doi.org/10.1038/s41598-020-68356-w.

Hill, S. A., 2019: Theories for past and future monsoon rainfall changes. *Current Climate Change Reports*, **5,** 160-171, https://doi.org/10.1007/s40641-019-00137-8.

Hirasawa, H., P. J. Kushner, M. Sigmond, J. Fyfe, and C. Deser, 2020: Anthropogenic Aerosols Dominate Forced Multidecadal Sahel Precipitation Change through Distinct Atmospheric and Oceanic Drivers. *J. Climate*, **33,** 10187-10204, https://doi.org/10.1175/JCLI-D-19-0829.1.

Hoerling, M., J. Hurrell, J. Eischeid, and A. Phillips, 2006: Detection and attribution of twentieth-century northern and southern African rainfall change. *J. Climate*, **19,** 3989-4008, https://doi.org/10.1175/JCLI3842.1.

Hua, W., A. Dai, L. Zhou, M. Qin, and H. Chen, 2019: An Externally Forced Decadal Rainfall Seesaw Pattern Over the Sahel and Southeast Amazon. *Geophys. Res. Lett.*, **46,** 923-932, https://doi.org/10.1029/2018GL081406.

Huang, B., and Coauthors, 2017: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Climate*, **30,** 8179-8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Hwang, Y. T., D. M. Frierson, and S. M. Kang, 2013: Anthropogenic sulfate aerosol and the southward shift of tropical precipitation in the late 20th century. *Geophys. Res. Lett.*, **40,** 2845-2850.

Iles, C. E., and G. C. Hegerl, 2014: The global precipitation response to volcanic eruptions in the CMIP5 models. *Environ. Res. Lett.*, **9**, https://doi.org/10.1088/1748-9326/9/10/104012.

Kang, S. M., D. M. Frierson, and I. M. Held, 2009: The tropical response to extratropical thermal forcing in an idealized GCM: The importance of radiative feedbacks and convective parameterization. *J. Atmos. Sci.*, **66,** 2812-2827, https://doi.org/10.1175/2009JAS2924.1.

Kang, S. M., I. M. Held, D. M. Frierson, and M. Zhao, 2008: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *J. Climate*, **21,** 3521-3532, https://doi.org/10.1175/2007JCLI2146.1.

Kawase, H., M. Abe, Y. Yamada, T. Takemura, T. Yokohata, and T. Nozawa, 2010: Physical mechanism of long-term drying trend over tropical North Africa. *Geophys. Res. Lett.*, **37**, https://doi.org/10.1029/2010GL043038.

Klimont, Z., S. J. Smith, and J. Cofala, 2013: The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. *Environ. Res. Lett.*, **8**, https://doi.org/10.1088/1748-9326/8/1/014003.

Knight, J. R., C. K. Folland, and A. A. Scaife, 2006: Climate impacts of the Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **33**, https://doi.org/10.1029/2006GL026242.

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.*, **32**, https://doi.org/10.1029/2005GL024233.

Knutson, T. R., and S. Manabe, 1995: Time-mean response over the tropical Pacific to increased C02 in a coupled ocean-atmosphere model. *J. Climate*, **8,** 2181-2199, https://doi.org/10.1175/1520-0442(1995)008<2181:TMROTT>2.0.CO;2.

Kucharski, F., N. Zeng, and E. Kalnay, 2013: A further assessment of vegetation feedback on decadal Sahel rainfall variability. *Climate Dynam.*, **40,** 1453-1466, DOI 10.1007/s00382-012-1397-x.

Losada, T., B. Rodriguez-Fonseca, E. Mohino, J. Bader, S. Janicot, and C. R. Mechoso, 2012: Tropical SST and Sahel rainfall: A non-stationary relationship. *Geophys. Res. Lett.*, **39**, https://doi.org/10.1029/2012GL052423.

Lu, J., 2009: The dynamics of the Indian Ocean sea surface temperature forcing of Sahel drought. *Climate Dynam.*, **33,** 445-460, https://doi.org/10.1007/s00382-009-0596-6.

Marvel, K., M. Biasutti, and C. Bonfils, 2020: Fingerprints of external forcings on Sahel rainfall: aerosols, greenhouse gases, and model-observation discrepancies. *Environ. Res. Lett.*, **15**, https://doi.org/10.1088/1748-9326/ab858e.

Menary, M. B., and Coauthors, 2020: Aerosol-forced AMOC changes in CMIP6 historical simulations. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2020GL088166.

Murphy, L. N., K. Bellomo, M. Cane, and A. Clement, 2017: The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **44,** 2472-2480, https://doi.org/10.1002/2016GL071337.

Neelin, J., C. Chou, and H. Su, 2003: Tropical drought regions in global warming and El Nino teleconnections. *Geophys. Res. Lett.*, **30**, https://doi.org/10.1029/2003GL018625.

Okonkwo, C., and Coauthors, 2015: Combined effect of El Niño southern oscillation and Atlantic multidecadal oscillation on Lake Chad level variability. *Cogent Geoscience*, **1**, https://doi.org/10.1080/23312041.2015.1117829.

Palmer, T., 1986: Influence of the Atlantic, Pacific and Indian oceans on Sahel rainfall. *Nature*, **322,** 251-253, https://doi.org/10.1038/322251a0.

Parhi, P., A. Giannini, P. Gentine, and U. Lall, 2016: Resolving contrasting regional rainfall responses to El Niño over tropical Africa. *J. Climate*, **29,** 1461-1476, https://doi.org/10.1175/JCLI-D-15-0071.1.

Park, J.-y., J. Bader, and D. Matei, 2016: Anthropogenic Mediterranean warming essential driver for present and future Sahel rainfall. *Nat. Clim. Change*, **6,** 941-945, https://doi.org/10.1038/nclimate3065.

Pearl, J., M. Glymour, and N. P. Jewell, 2016: *Causal inference in statistics: A primer.* John Wiley & Sons, 136 pp.

Polson, D., M. Bollasina, G. Hegerl, and L. Wilcox, 2014: Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols. *Geophys. Res. Lett.*, **41,** 6023-6029, https://doi.org/10.1002/2014GL060811.

Pomposi, C., Y. Kushnir, and A. Giannini, 2015: Moisture budget analysis of SST-driven decadal Sahel precipitation variability in the twentieth century. *Climate Dynam.*, **44,** 3303-3321, https://doi.org/10.1007/s00382-014-2382-3.

Pomposi, C., A. Giannini, Y. Kushnir, and D. E. Lee, 2016: Understanding Pacific Ocean influence on interannual precipitation variability in the Sahel. *Geophys. Res. Lett.*, **43,** 9234-9242, https://doi.org/10.1002/2016GL069980.

Qin, M., A. Dai, and W. Hua, 2020: Quantifying contributions of internal variability and external forcing to Atlantic multidecadal variability since 1870. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2020GL089504.

Rahmstorf, S., J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, and E. J. Schaffernicht, 2015: Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat. Clim. Change*, **5,** 475-480, https://doi.org/10.1038/nclimate2554.

Rodríguez-Fonseca, B., and Coauthors, 2015: Variability and predictability of West African droughts: A review on the role of sea surface temperature anomalies. *J. Climate*, **28,** 4034-4060, https://doi.org/10.1175/JCLI-D-14-00130.1.

Rosenfeld, D., and Coauthors, 2008: Flood or drought: How do aerosols affect precipitation? *Science*, **321,** 1309-1313, https://doi.org/10.1126/science.1160606.

Rotstayn, L. D., and U. Lohmann, 2002: Tropical rainfall trends and the indirect aerosol effect. *J. Climate*, **15,** 2103-2116, https://doi.org/10.1175/1520-0442(2002)015<2103:TRTATI>2.0.CO;2.

Scaife, A., and Coauthors, 2009: The CLIVAR C20C project: selected twentieth century climate events. *Climate Dynam.*, **33,** 603-614, https://doi.org/10.1007/s00382-008-0451-1.

Schneider, T., T. Bischoff, and G. H. Haug, 2014: Migrations and dynamics of the intertropical convergence zone. *Nature*, **513,** 45-53, https://doi.org/10.1038/nature13636.

Smith, S. J., J. v. Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias, 2011: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.*, **11,** 1101-1116, https://doi.org/10.5194/acp-11-1101-2011.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate*, **15,** 2702-2706, https://doi.org/10.1175/1520-0442(2002)015<2702:TESITT>2.0.CO;2.

Sutton, R. T., and D. L. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309,** 115-118, https://doi.org/10.1126/science.1109496.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.*, **93,** 485-498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, **22,** 1469-1481, https://doi.org/10.1175/2008JCLI2561.1.

Undorf, S., D. Polson, M. Bollasina, Y. Ming, A. Schurer, and G. Hegerl, 2018: Detectable impact of local and remote anthropogenic aerosols on the 20th century changes of West African and South Asian monsoon precipitation. *J. Geophys. Res.-Atmos.*, **123,** 4871-4889, https://doi.org/10.1029/2017JD027711.

Vellinga, M., and Coauthors, 2016: Sahel decadal rainfall variability and the role of model horizontal resolution. *Geophys. Res. Lett.*, **43,** 326-333, https://doi.org/10.1002/2015GL066690.

Westervelt, D., and Coauthors, 2017: Multimodel precipitation responses to removal of US sulfur dioxide emissions. *J. Geophys. Res.-Atmos.*, **122,** 5024-5038, https://doi.org/10.1002/2017JD026756.

Yan, X., R. Zhang, and T. R. Knutson, 2018: Underestimated AMOC Variability and Implications for AMV and Predictability in CMIP Models. *Geophys. Res. Lett.*, **45,** 4319-4328, https://doi.org/10.1029/2018GL077378.

——, 2019: A multivariate AMV index and associated discrepancies between observed and CMIP5 externally forced AMV. *Geophys. Res. Lett.*, **46,** 4421-4431, https://doi.org/10.1029/2019GL082787.

Zhang, R., 2017: On the persistence and coherence of subpolar sea surface temperature and salinity anomalies associated with the Atlantic multidecadal variability. *Geophys. Res. Lett.*, **44,** 7865-7875, https://doi.org/10.1002/2017GL074342.

Zhang, R., and T. L. Delworth, 2006: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, https://doi.org/10.1029/2006GL026267.

Zhang, R., R. Sutton, G. Danabasoglu, T. L. Delworth, W. M. Kim, J. Robson, and S. G. Yeager, 2016: Comment on "The Atlantic Multidecadal Oscillation without a role for ocean circulation". *Science*, **352,** 1527-1527, https://doi.org/10.1126/science.aaf1660.

Zhang, R., and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.*, **70,** 1135-1144, https://doi.org/10.1175/JAS-D-12-0331.1.