Constraining the pattern and magnitude of projected extreme precipitation change in a multi-model ensemble

Maximilian Kotz[a b], Stefan Lange [a], Leonie Wenz [a c], Anders Levermann [a b d]

[a] *Potsdam Institute for Climate Impact Research*

[b] *Innstitute of Physics, University of Potsdam*

[c] *Mercator Research Institute on Global Commons and Climate Change*

[d] *Lamont-Doherty Earth Observatory, Columbia University*

Corresponding author: Maximilian Kotz, maxkotz@pik-potsdam.de

Early Online Release: This preliminary version has been accepted for publication in *Journal of Climate*, may be fully cited, and has been assigned DOI 10.1175/JCLI-D-23-0492.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT: Projections of precipitation extremes over land are crucial for socioeconomic risk assessments, yet model discrepancies limit their application. Here we use a pattern-filtering technique to identify low-frequency changes in individual members of a multi-model ensemble, to assess discrepancies across models in the projected pattern and magnitude of change. Specifically, we apply low-frequency component analysis (LFCA) to the intensity and frequency of daily precipitation extremes over land in 21 CMIP-6 models. LFCA brings modest but statistically significant improvements in the agreement between models in the spatial pattern of projected change, particularly in scenarios with weak greenhouse forcing. Moreover, we show that LFCA facilitates a robust identification of the rates at which increasing precipitation extremes scale with global temperature change within individual ensemble members. While these rates approximately match expectations from the Clausius-Clapeyron relation on average across models, individual models exhibit considerable and significant differences. Monte-Carlo simulations indicate that these differences contribute to uncertainty in the magnitude of projected change at least as much as differences in the climate sensitivity. Finally, we compare these scaling rates to those identified from observational products, demonstrating that virtually all climate models significantly underestimate the rates at which increases in precipitation extremes have scaled with global temperatures historically. Constraining projections with observations therefore amplifies the projected intensification of precipitation extremes as well as reducing the relative error of their distribution.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

1. Introduction

An intensification of the hydrological cycle (Ziegler et al. (2003)) is likely to play a major role in the socio-economic impacts of climate change. Key determinants of social welfare, such as groundwater availability (Thomas and Famiglietti (2019)), agricultural productivity (Liang et al. (2017)), and social stability (Hsiang et al. (2013); von Uexkull et al. (2016)) are closely linked to changes in precipitation. In particular, the changing frequency and intensity of heavy precipitation extremes constitutes an important channel, given their links to flooding which can cause considerable loss of life (Schumacher (2017)), financial losses (Davenport et al. (2021); Frame et al. (2020)), as well as impacts on overall macroeconomic productivity (Kotz et al. (2022)). Understanding how these characteristics of precipitation will change under anthropogenic influence is therefore crucial for informing risk assessments and climate policy. Climate models such as those in the Coupled Model Intercomparison Project (CMIP6; Eyring et al. (2016)) play a crucial role in this understanding, providing projections of precipitation extremes under different levels of greenhouse forcing at the regional and temporal detail necessary for assessment of these impacts (Warszawski et al. (2014)). These projections can subsequently inform detailed assessments of the consequences of both mitigation (Lange et al. (2020); Thiery et al. (2021)) and adaptation (Willner et al. (2018); Boulange et al. (2021)).

Despite considerable progress, barriers to an effective use of these projections remain due to uncertainty in projected precipitation change (Knutti and Sedláček (2013); Shepherd (2014)). The simplest theory regarding precipitation extremes indicates that the scaling of available moisture with atmospheric temperatures based on the Clausius-Clapeyron relation (CC-relation) (Allen and Ingram (2002); Santer et al. (2007); Fischer and Knutti (2016)) should lead to increases in precipitation extremes with global warming at rates of 6-7$\%K^{-1}$. Indeed, intensification of precipitation extremes have been detected in almost all global land areas (Min et al. (2011); Zhang et al. (2013); Fischer and Knutti (2016); Chen and Sun (2017); Kirchmeier-Young and Zhang (2020); Madakumbura et al. (2021)), and when aggregated spatially are largely consistent with expectations from the CC-relation (Fischer and Knutti (2016)). However, a number of other factors determine changes in precipitation extremes including the vertical structure of upwards atmospheric velocities and their change (O'Gorman and Schneider (2009)), as well as shifts in atmospheric circulation (Pfahl et al. (2017)). These factors contribute significantly to regional differences in

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

the intensification of precipitation extremes (O'Gorman and Schneider (2009)), and are also the dominant source of uncertainty across climate model projections (Pfahl et al. (2017)).

In addition to these uncertainties in the response of precipitation extremes to human forcing, further uncertainty in future projections exists due to the role of chaotic internal variability (Shepherd (2014); Fischer et al. (2014)). This is particularly relevant when internal variability is large in comparison to the strength of the anthropogenically-forced signal. For example, when considering changes over short timescales, in scenarios of weak greenhouse forcing or in variables with weaker response to anthropogenic influence (Blanusa et al. (2023)). The latter is certainly true for changes in precipitation compared to temperature (King et al. (2015)), emphasising the possibility that internal variability may mask or bias assessments of how precipitation extremes change in response to human influence. An effective method to quantify and separate the contributions of human forcing and internal variability to projected precipitation change is the use of large ensembles of climate models initialised from different starting conditions (Aalbers et al. (2018); Wood and Ludwig (2020)). However, such Single Model Initial-condition Large Ensembles (SMILEs) (Kay et al. (2015)) do not enable the investigation of discrepancies which may exist between different climate models in the simulated response of precipitation extremes to human forcing.

Here we aim to assess uncertainties in projected changes in daily precipitation extremes which arise from the different responses to anthropogenic forcing embodied in the different members of the multi-model ensemble CMIP-6. Our interest in the differences between climate models constrains us to identify changes in each ensemble member individually. To limit the influence of chaotic internal variability, we therefore use a pattern-filtering methodology which has been designed to separate the forced response of the climate system from internal variability within individual ensemble members. This method, known as low-frequency component analysis (LFCA), extends an Empirical Orthogonal Function (EOF) decomposition to search for linear-re combinations of the orthogonal basis functions with a maximal ratio of low-frequency to total variance (Wills et al. (2018)). Given the lower-frequency at which anthropogenically forced signals evolve in comparison to chaotic internal variability, this approach can help to discriminate between them while drawing on the structure of the spatiotemporal covariance matrix to do so. When tested in the context of large ensembles, this method shows particular promise in separating the anthropogenic component of projected temperature change from individual ensemble members with greater accuracy than

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

large ensembles with up to 20-members (Wills et al. (2020)). Although not yet tested against large ensembles in the context of precipitation, this method nevertheless presents an opportunity to limit the role of internal variability when assessing discrepancies in the projected changes of extreme precipitation in individual members of the multi-model CMIP6 ensemble.

We assess changes in both the intensity and frequency of daily precipitation extremes, focusing on changes over land because of their relevance for impacts (Davenport et al. (2021); Kotz et al. (2022)), and for better comparison to observations as is made in the final section. In the following Section 2 we provide an overview of the data and methods. Section 3a presents the low-frequency changes in extreme precipitation detected by LFCA in the individual members of the CMIP6 ensemble. In section 3b we assess discrepancies across models in the spatial patterns of projected change and the extent to which LFCA limits this uncertainty. Section 3c addresses how differences in the magnitude of projected precipitation change are constrained by scaling relations with global temperatures between climate models, but also within individual models. Finally, in Section 3d we assess the contribution of differences in temperature-precipitation scaling rates to uncertainty in the magnitude of projected change, as well as how these rates compare to those observed historically.

2. Data and Methods

a. Climate data

We use daily surface precipitation rates and daily 2-m temperature from 21 climate models participating in CMIP6 (Eyring et al. (2016)). We choose models which provide output under both the historical (1850-2014) and the future (2015-2100) greenhouse forcing scenarios specified by SSP126 and SSP585. Since our approach is primarily interested in differences between climate models, and since many models provide only a single ensemble member, we use only the first ensemble member available for each model. A full list of models is displayed in Fig. 5 and Supplementary Figs. S1-4. Additionally, we use daily 2-m temperature and daily precipitation totals from ERA5 (Hersbach et al. (2020)), as well as daily precipitation totals from the Global Precipitation Climatology Centre (Schneider et al. (2008)).

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

b. Extreme precipitation metrics

We assess changes in both the intensity and frequency of daily precipitation extremes, following work indicating that both are relevant for impacts (Kotz et al. (2022)). We assess the intensity of daily extremes by taking the annual maximum, RX1, of daily precipitation, $P_{x,d}$:

$$RX1_{x,y} = \max(P_{x,d} : x = 1, ...N_y). \tag{1}$$

The annual frequency of daily extremes is assessed by counting the number of days in each year which exceed the 99^{th} percentile of the historical distribution:

$$R > 99\%_{x,y} = \sum_{d=1}^{N_y} H(P_{x,d} > 99\%) \tag{2}$$

where N_y is the number of days in a given year, H is the Heaviside step function, and x and d respectively denote grid-cell and day. The historical period used to estimate percentiles is either the pre-industrial period (1850-1950), or the historical period over which observational data are available (1985-2015), depending on the use. These annual measures are estimated on the native CMIP grids, before being linearly interpolated to a common 1-by-1 degree grid for further analysis.

c. Low-frequency component analysis

We use LFCA to estimate low-frequency changes in extreme precipitation. LFCA uses an EOF decomposition and subsequent linear-recombination under a frequency-dependent variance maximisation constraint to optimally filter low-frequency changes from higher-frequency modes. As such, it has been designed (Wills et al. (2018)) and demonstrated (Wills et al. (2020)) to separate forced climate changes from chaotic internal variability. We use LFCA over other methods, such as signal-to-noise maximising pattern-filtering, because of our interest in identifying changes in individual members of the multi-model CMIP-6 ensemble, in which context LFCA provides a superior performance (Wills et al. (2020)). Here we provide a conceptual summary of LFCA and of its application to identifying the climatic response to anthropogenic forcing, please see Wills et al. (2018) for a more detailed introduction to and description of the method.

Linear recombinations of the leading empirical-orthogonal-functions (EOFs) are found which maximise the ratio of low-frequency to total variance which they can explain. We retain a sufficient

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

number of EOFs to account for a minimum of 70% of the original spatio-temporal variance, and define low-frequency variance as that resulting from applying a 20-year low-pass Butterworth filter with reflecting boundary conditions to grid-cell level departures from linear trends. We use a lower cut-off frequency than Wills et al. (2018, 2020), due to the lower signal-to-noise ratio of the climate change signal in precipitation than temperature (Deser et al. (2014)), but recover consistent results under alternative filtering specifications. The resulting linear recombinations are independent, and ordered in terms of increasing frequency. They constitute both a "low-frequency component" (LFC) and "low-frequency pattern" (LFP); the LFC is a time-series which describes the temporal evolution of the specific spatial pattern encompassed by the LFP. The lowest-frequency component(s) of LFCA have been found to accurately characterise the forced response of the climate system to anthropogenic forcing in the context of temperature (Wills et al. (2020)) and temperature variability (Kotz et al. (2021)). However, given that tests of such methods in large ensembles have not yet been conducted for precipitation extremes, we here interpret the lowest-frequency component identified by LFCA as the lowest-frequency change projected by each ensemble member rather than the forced change.

We apply LFCA to annual time series of the intensity (Rx1) and frequency (R>99%) of daily precipitation extremes over land-areas and to annual mean temperature (over the globe) from 1950-2100 under the anthropogenic forcing of the historical period and both the SSP126 and SSP585 future scenarios, having first linearly interpolated each index to a common 1-by-1 degree grid. Low-frequency changes between two given time periods (usually between two decades) are then calculated as the product of the lowest-frequency LFP and the difference between temporal (usually decadal) averages of the corresponding LFC. In a number of cases, we also estimate the forced changes in each precipitation metric using temporal averages of the data without the use of LFCA to first identify low-frequency changes. In these cases we either use 10, 20 or 30 year periods to estimate the change.

d. Assessing similarity in the patterns of change

We use two metrics to assess similarities in the patterns of change detected by LFCA. The first is simply the number of ensemble members which agree on the projected sign of change in a given region. This metric is aggregated in Fig. 2 over global land areas to indicate the fraction of

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

land-area on which a certain percentage of models are in agreement. The second metric is a centred pattern correlation as defined in Santer et al. (1995). This metric reflects differences in the relative magnitude of change as well as the sign of change. Pattern correlations are estimated between each unique pair of climate models (210 pairs) and distributions of these pattern correlations are shown and used to estimate the significance of improvements in inter-model agreement.

e. Scaling between precipitation extremes and global temperature change

Spatially-averaged changes in precipitation extremes are compared to changes in global temperature change. Changes in global temperature are calculated by taking area-weighted averages of changes in global 2-m surface air temperature (SAT) estimated from the lowest-frequency component identified with LFCA. Changes in precipitation extremes are expressed as local percentages before taking an area-weighted average, to make comparison to the theoretical expectations of the CC-relation. We explore the sensitivity of our results to arid-regions with low baseline values in latter parts of the manuscript. Changes are estimated either over the whole simulation period (1950-60 to 2090-2100, as in Fig. 3), or between pairs of non-overlapping decades separated by 30-years within each individual model (Fig. 4 on-wards). In the first case, percentages are expressed in relation to the pre-industrial baseline (1850-1950), land-area-averages are taken, and scaling relations are then estimated in a logarithmic form, $log(\frac{\delta P}{P} + 1)$ to reflect the exponential scaling anticipated from the CC-relation. In the second case, percentages are estimated with respect to the baseline periods from which each change is estimated. Here, since the baseline period for estimating percentage changes is updated with each differencing-period, only a linear-scaling is estimated, after taking land-area averages of local percentage changes.

Temperature-precipitation scaling relations are estimated using least-squares linear regressions. In each case, the intercept of these regressions is set to zero to reflect the physical constraint that without changes in global SAT there should be no change in precipitation extremes. Methodological uncertainty in the estimates of these scaling rates is obtained via a bootstrapping approach. In Fig. 3, the climate models are re-sampled 1000 times with replacement, whereas in Fig. 4 the different time periods are re-sampled equivalently. As such, these estimates reflect the methodological uncertainty arising from the availability of climate models (in Fig. 3) and the temporal evolution of low-frequency changes in projected precipitation change (in Fig. 4). When estimating the

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

significance of differences between model scaling rates, we estimate a distribution of differences between scaling rates from the estimates obtained from the 1000 resamples of the time series, and assess whether the 95% confidence intervals of this distribution of differences encompass zero.

f. Monte-Carlo simulations to propagate uncertainties and assess relative contributions

We use a Monte-Carlo simulation procedure to assess how different sources contribute to uncertainty in the overall magnitude of projected extreme precipitation change. Specifically, we assess two contributions, from a) differences in the projected change in global mean temperatures for a given forcing scenario (loosely equivalent to the transient climate sensitivity), and b) differences in the rates at which precipitation extremes scale with global temperature change. We sample 10,000 times from the distribution of these two parameters and combine them to provide estimates of the distribution of projected change under uncertainty from both sources. This approach is justified by the fact that temperature-precipitation scaling rates appear independent of global temperature change across models. The distribution of projected global temperature change is obtained from the 21 different climate models of the ensemble, while the distribution of temperature-precipitation scaling rates comes from both the different climate models and the bootstrapped assessments of uncertainty in their individual scaling rates (as described above). Finally, we estimate a simple observational-constraint by replacing the distribution of temperature-precipitation scaling rates with the estimates obtained from the two observational data products.

3. Results

a. Low-frequency changes in projected precipitation extremes

We identify low-frequency changes in the intensity and frequency of daily precipitation extremes under historical and future (SSP585) anthropogenic forcing using LFCA. Results shown in Fig. 1 are for a selection of ten of the twenty-one CMIP-6 models, taking every other model when ordered in terms of SAT change (see Supplementary Figs. S1-4 for results from all members of the multi-model ensemble under both SSP585 and SSP126 future forcing). For each precipitation metric and for each model, the lowest-frequency component (LFC-1) exhibits a near-monotonic trend which closely follows the increasing concentrations of greenhouse gases in the historical and SSP585 scenario. However, both the intensity and pattern of low-frequency change (1950-60 to

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

2090-2100) show clear differences between models. While most land-areas exhibit increases in both the intensity and frequency of extremes, stronger increases are typically found across tropical regions. Decreases are projected in certain models, in particular in regions where net precipitation is projected to decline such as the Mediterranean basin, parts of Australia, South Africa and central South America (Masson-Delmotte et al. (2021)). These regions also correspond to those identified by Pfahl et al. (2017) as having larger contributions from changes in atmospheric circulation.

Comparing the projected low-frequency changes in extreme precipitation to the magnitude of historical variability (standard deviation of annual time series from 1850-1900), we find that changes by the end of the century are not always significant with respect to the magnitude of internal variability (see stippled regions in Fig. 1). Changes are most often of a significant magnitude across the tropics and high-latitudes, showing the least significance across mid-latitudes and in regions where precipitation extremes are projected to decrease. Moreover, models with larger SAT change typically show significant changes over larger portions of the global land-area (Fig. A1a-d). This reflects the fact that SAT change is a strong determinant of the magnitude of changes in precipitation extremes (explored further in section 3c), and therefore of the signal-to-noise ratio of forced change to internal variability. This can be further seen in the fact that models with larger SAT change contain a larger portion of the overall spatio-temporal variance within the lowest-frequency component detected by LFCA (Fig. A1e-h). These results are robust to methodological choices regarding the LFCA, such as the specific percentage of spatio-temporal variance which is retained following the EOF decomposition as well as the time-scale of the filter used to define low-frequency variance.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

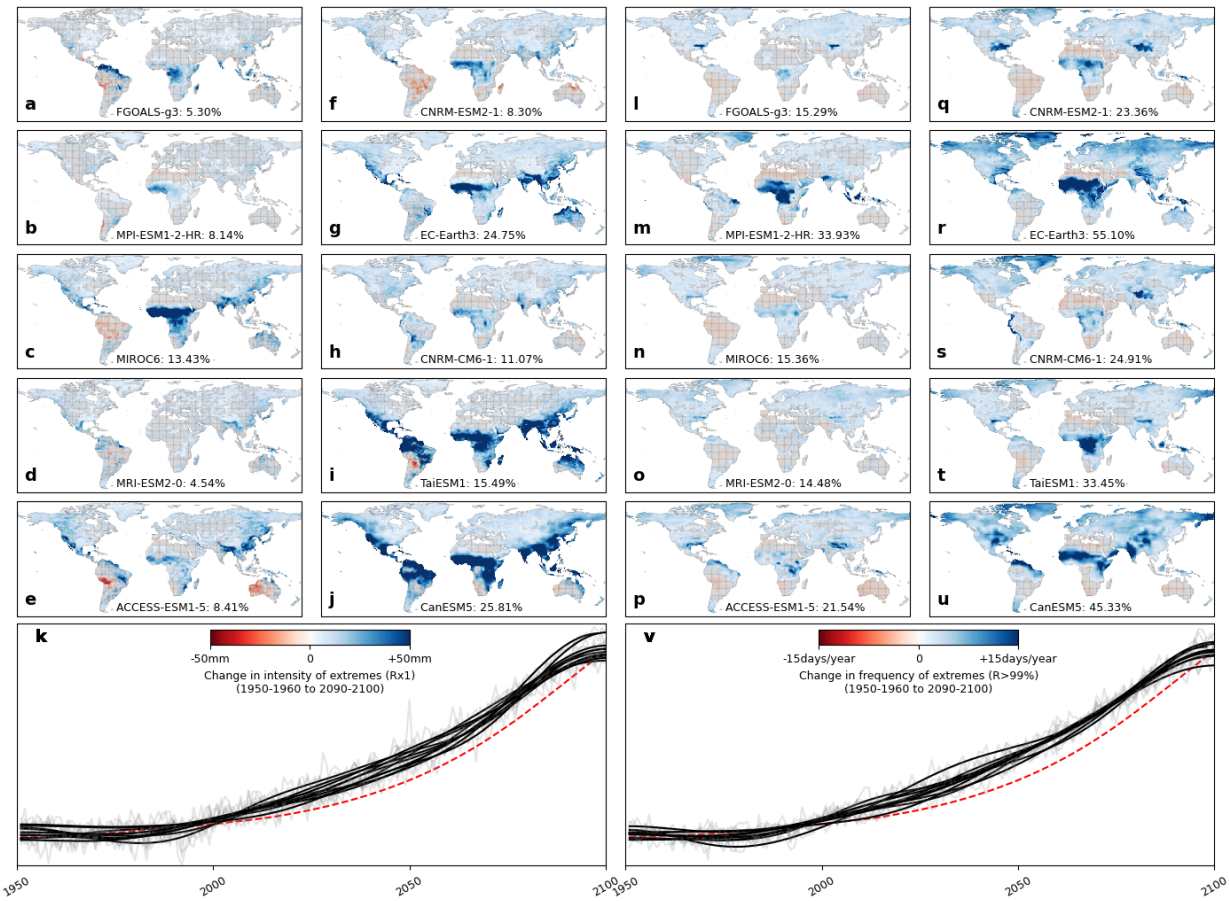

FIG. 1. Low-frequency changes in extreme daily precipitation under historical (1950-2014) and future (SSP585, 2015-2100) anthropogenic forcing, detected in individual CMIP6 climate models with low-frequency component analysis. (a-j) The spatial pattern of low-frequency change in the intensity of daily precipitation extremes (annual maximum, Rx1) from 1950-60 to 2090-2100 (the product of the lowest-frequency pattern with the difference between decadal averages of its corresponding component) for 10 of the 21 CMIP6 ensemble members. Regions are stippled where the projected change is smaller than the historical variability, estimated as the standard deviation of annual values over the period 1850-1900. (k) The temporal evolution of the corresponding lowest-frequency components (LFC-1) are shown in grey with a 20-year Butterworth filtered time-series in black. Time series for each model are overlain due to their similarity. The concentration of greenhouse gases in the historical period and SSP585 are rescaled and shown in red for comparison. (l-v) Corresponding results for changes in the frequency of daily precipitation extremes (the number of days with rainfall greater than the 99th percentile over the historical period 1850-1950). The model name is indicated in the bottom of each panel, along with the percentage of total spatio-temporal variance accounted for by the lowest-frequency component. We here show models chosen evenly from across the distribution of modelled global mean surface air temperature (SAT) change (ordered by increasing SAT change), and show results for all ensemble members for each precipitation metric under SSP585 and SSP126 forcing scenarios in Supplementary Figs. S1-S4

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

b. Discrepancies in the spatial pattern of projected low-frequency change

Having identified low-frequency changes within each member of the multi-model ensemble, we assess discrepancies in the spatial pattern of projected change between models (Fig. 2). In the ensemble mean, projected low-frequency changes reflect a near-global intensification of the intensity and frequency of daily precipitation extremes (Fig. 2a, b), with some weak decreases in regions where reductions in net precipitation are particularly strong (Masson-Delmotte et al. (2021)) and where changes in atmospheric circulation contribute more considerably (Pfahl et al. (2017)). Agreement between models is fairly high, with 90% of models agreeing on the sign of change on approximately 80% and 70% of the global land-area for changes in the intensity and frequency of extremes respectively in the high forcing scenario (Fig. 2d, f). Discrepancies are concentrated across the Mediterranean basin, Australia, South Africa and Central South America (Fig. 2a,b). These regions are those where decreases in precipitation extremes are projected, where projected changes are less significant with respect to internal variability, and where atmospheric circulation has been identified as a dominant source of uncertainty (Pfahl et al. (2017)).

Agreement across climate models on the spatial pattern of change is smaller under the scenario with weaker greenhouse forcing (SSP126). This is clear both in terms of the fraction of land-area on which models project the same sign of change (Fig. 2c-f) and also in terms of the distribution of pair-wise centred pattern correlations between climate models (Fig. 2g-j). This likely reflects the smaller signal-to-noise ratio of anthropogenically forced precipitation change compared to chaotic internal variability in the weaker forcing scenario. This can be seen in the smaller magnitude of projected precipitation change and the smaller percentage of spatiotemporal variance explained by the lowest-frequency component in the weaker forcing scenario (Fig. A1e-h). When internal variability is relatively larger, it will mask the anthropogenically forced signal identified by LFCA to a greater extent. This will lead to greater disagreement between models resulting from both discrepancies in the forced response, and the larger differences due to internal variability.

Furthermore, the extent of agreement between models is higher when using LFCA to identify low-frequency changes in precipitation extremes rather than temporal averages over 10, 20 and 30 year periods. These benefits manifest both in terms of the fraction of land-area in agreement (Fig. 2c-f) and the centred pattern correlations (Fig. 2g-j). Improvements are strongest in the weaker forcing scenario and when compared to taking averages over shorter periods. Under SSP126

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

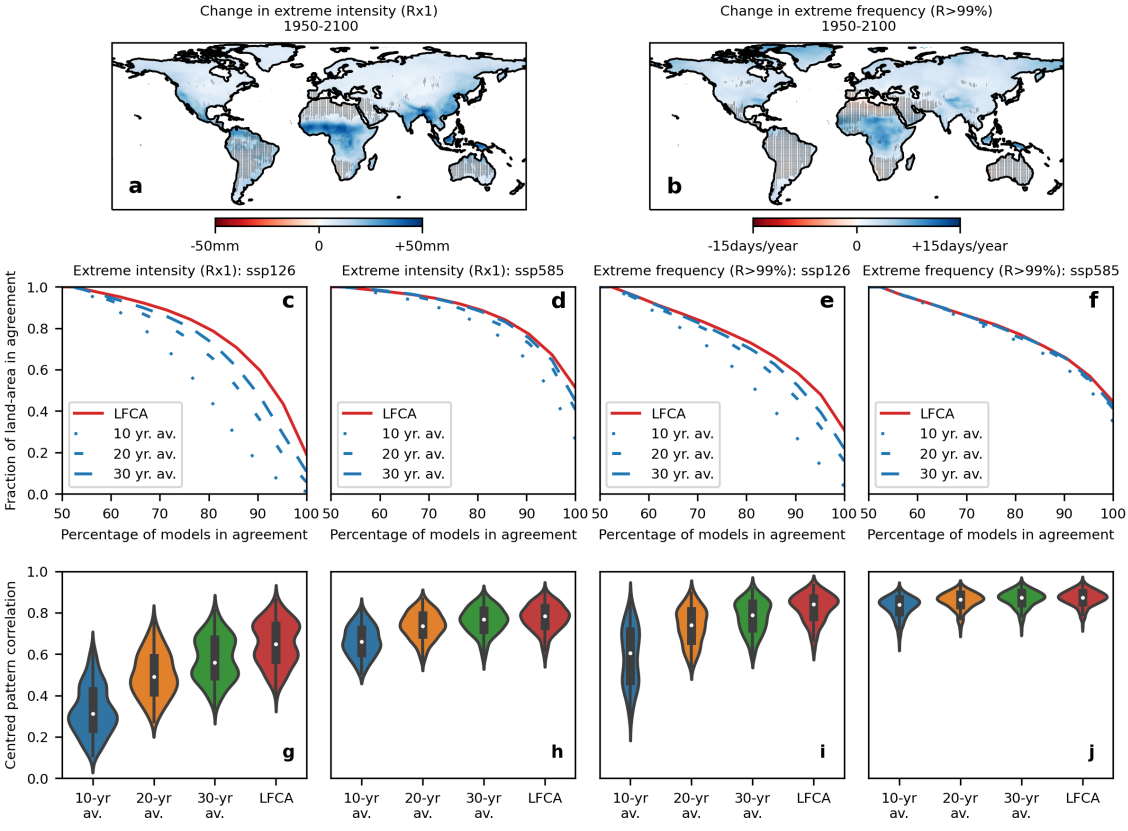

FIG. 2. Assessing discrepancies in the spatial pattern of low-frequency change in extreme precipitation. (a-b) The CMIP-6 ensemble-mean low-frequency change in the intensity (a) and frequency (b) of daily precipitation extremes. Changes are estimated from the lowest-frequency component detected with LFCA (see Fig. 1) over the period 1950-2100 under historical and SSP585 future anthropogenic forcing. Regions are stippled where less than 90% of models agree on the sign of projected change. (c-f) The fraction of land-area on which a certain percentage of CMIP-6 models agree on the sign of low-frequency change for the intensity (c-d) and frequency (e-f) of daily precipitation extremes under the low- (c, e) and high-greenhouse forcing scenarios (d, f), SSP126 and SSP585. Results are shown when having used either LFCA (red) or averages over 10, 20 and 30 years (blue) to detect low-frequency changes. (g-j) Distributions of centred pattern correlations between the low-frequency change projected by unique pairs of models for the two extreme precipitation metrics under the two forcing scenarios, when using the different methods to detect low-frequency changes. Violin plots from the Seaborn Python package indicate the inter-quartile range, upper and lower limits, as well as a Kernel-density distribution.

forcing, improvements in centred pattern correlations are found in 99 and 93% of model pairs when compared even to using 30-year averages, and differences between the distributions of correlations

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

obtained using the two methods are significant at the 0.1% level. Improvements in the stronger forcing scenario for the intensity of precipitation extremes are still significant at the 5% level, with 88% of model pairs showing improved agreement (Table 1). Assuming that multi-model agreement is an indicator of the extent to which the anthropogenic signal has been separated from chaotic internal variability, as outlined above, these improvements suggest that LFCA may improve the detection of the response of precipitation extremes to anthropogenic forcing. This conclusion would require further verification in the context of SMILEs which is beyond the scope of this manuscript. Nevertheless, these results demonstrate that LFCA significantly improves agreement in the pattern of projected change between members of a multi-model ensemble, particularly in the context of weak anthropogenic forcing.

TABLE 1. Assessing the significance of improvements in multi-model agreement on the spatial pattern of low-frequency change in extreme precipitation metrics. The percentage of unique model pairs with improved centred pattern correlations in low-frequency precipitation change when detected using LFCA compared to averages over 10, 20 and 30 year periods. The significance of the difference between the distributions of pattern correlations between models under the two different methods is assessed using the non-parametric Mann-Whitney test, with significance at the 5, 1, and 0.1% indicated by *, ** and ***.

Metric: SSP	LFCA>10-yr av.	LFCA>20-yr av.	LFCA>30-yr av.
Rx1: ssp126	100***	100***	99***
Rx1: ssp585	100***	100***	88*
R>99%: ssp126	100***	95***	93***
R>99%: ssp585	97***	82**	58

c. The scaling of precipitation extremes with global temperatures

In addition to uncertainty in the pattern of projected low-frequency change, the magnitude of extreme precipitation changes also exhibit large differences across models (Fig. 1). Changes in global SAT explain a large proportion of these differences when averaged over global land-areas (r^2 of 0.73 and 0.74 for the intensity and frequency of extremes, Fig. 3). Moreover, changes in the intensity of extremes scale with global SAT at a rate of $6.5\%K^{-1}$, closely following the expectations of the CC-relation within the uncertainty of the scaling relation we estimate from the climate models. For the frequency of extremes, we assess a theoretical expectation from the CC-relation by scaling up every day of the historical distribution of precipitation by $7\%K^{-1}$ and re-estimating

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

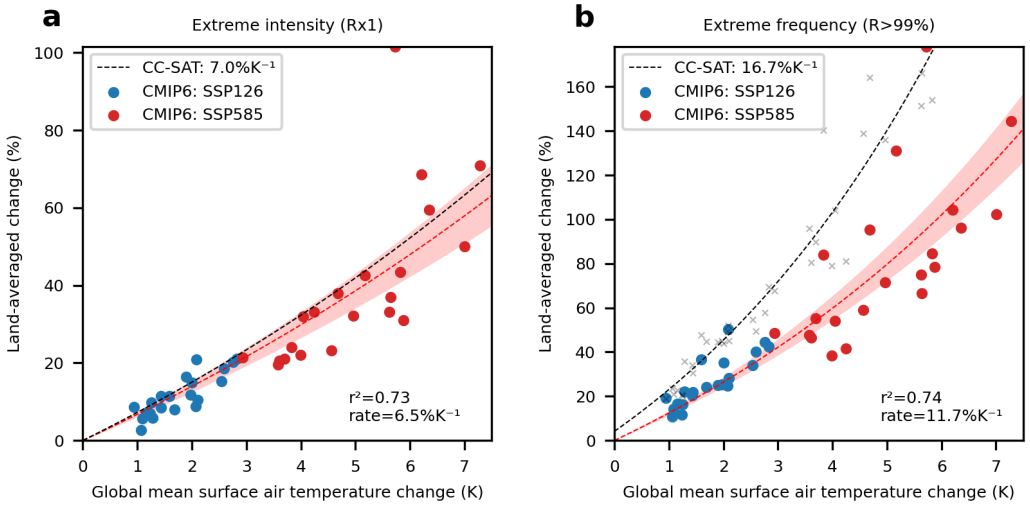

FIG. 3. The scaling of precipitation extremes over land with global mean surface air temperature (SAT) across CMIP6 models and scenarios. Low-frequency changes between 1950-1960 and 2090-2100 are calculated from the lowest-frequency component of each precipitation metric (as in Fig. 1) and of annual mean temperature. Red and blue colors denote the SSP585 and SSP126 scenarios of future greenhouse forcing. The results of least-squares linear regressions are shown in dashed red with the 5^{th} and 95^{th} confidence intervals shaded based on bootstrapped estimates of the regression (1000 climate model resamples with replacement). Changes in extreme metrics are calculated as a percentage of the historical baseline (1850-1950) for comparison to the theoretical expectations of the Clausius-Clapeyron (CC) relation, shown in dashed black lines. For the frequency of daily extremes (P>99p), a theoretical CC relation is estimated by scaling up each day of the historical precipitation distribution (1850-1950) by the given level of SAT change, and re-calculating each index, following Fischer and Knutti (2016). Individual estimates from this method are shown in grey, the black dashed-line showing the result of an exponential regression to these estimates. The resulting theoretical scaling rate of this regression is displayed in the figure legend.

the excedence of the original percentile-based thresholds (following the methods of Fischer and Knutti (2016)). This theoretical CC-relation predicts a scaling rate of $16.7\%K^{-1}$, and we find that the frequency of daily extremes do indeed increase at a higher rate of $11.7\%K^{-1}$, but one which falls considerably below this theoretical expectation. These results replicate many recent findings that changes in precipitation extremes scale with global SAT change across climate models, emphasising the importance of limiting global warming to limit future impacts as well as the importance of the climate sensitivity as a dominant contributor to their uncertainty. Furthermore, results are robust

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

to excluding regions with very low initial precipitation values which could potentially bias our assessment using land-area averages of local percentage changes (see Supplementary Fig. S5).

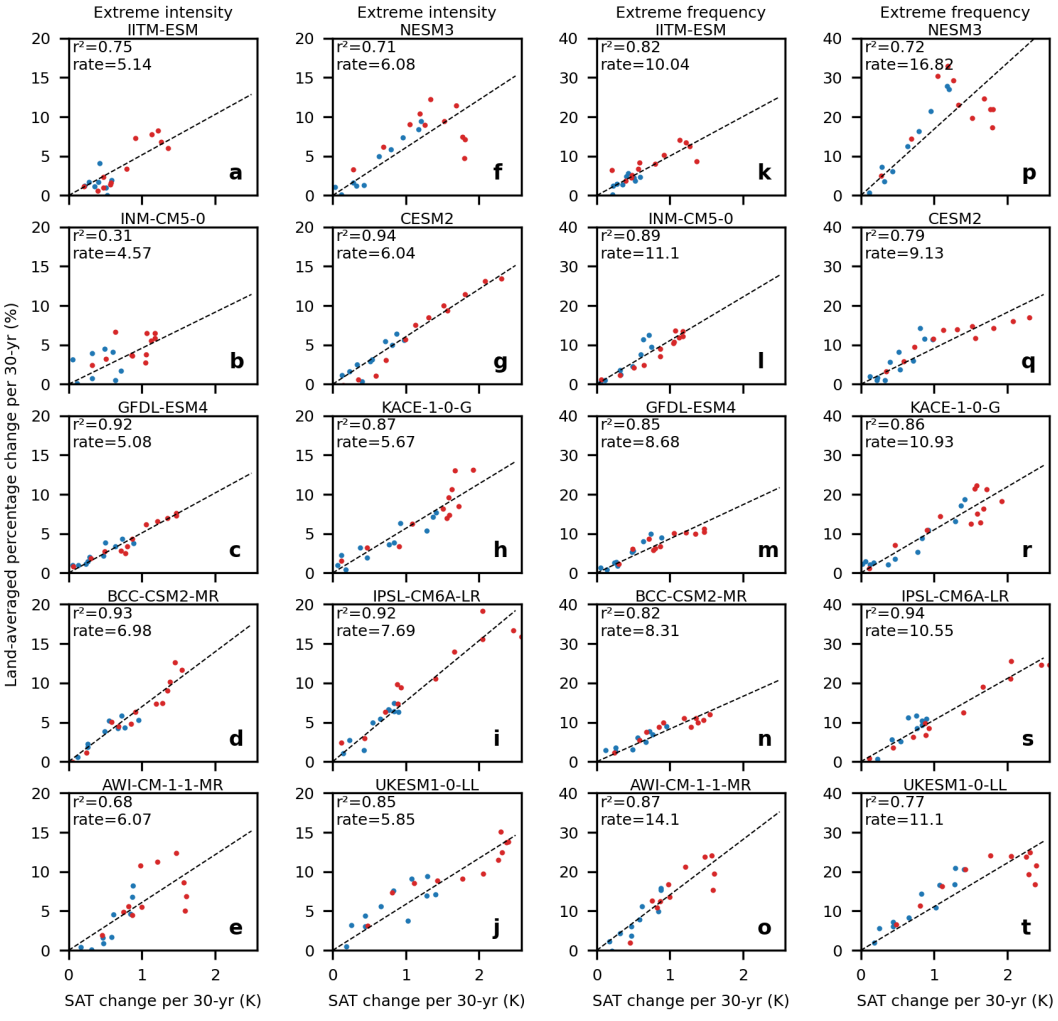

FIG. 4. The scaling of precipitation extremes over land with global mean surface air temperature (SAT) within individual CMIP6 climate models. (a-j) Land-averaged percentage changes in the lowest-frequency component of extreme intensity (Rx1) between pairs of non-overlapping decades separated by 30 years under a low- (SSP126, blue) and high-greenhouse forcing scenario (SSP585, red). Note that percentage changes are here estimated with respect to the first of the two decades between which changes are estimated, in contrast to Fig. 3 where they are estimated with respect to the baseline period 1850-1950. The results of a least-squares regression are shown in black. (k-t) Equivalent results for changes in extreme frequency. 10 of the 21 CMIP-6 models are shown, ordered from lowest to highest SAT change, while the full ensemble are shown in Figs. B2-3.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

While the results of Fig. 3 demonstrate that the intensification of precipitation extremes is largely determined by SAT changes, it is of further interest whether such temperature-precipitation scaling relations can be identified within individual models, and if so, how their rates may differ. We address this question by using the inter-temporal changes in precipitation extremes and global SAT within a given ensemble member, taking the changes occurring between non-overlapping decades separated by 30 year periods within each of the two forcing scenarios for each climate model. This approach reveals robust temperature-precipitation scaling relations which explain a large proportion of the inter-temporal changes in precipitation within each individual ensemble member (Fig. 4). The average R^2 of these relations is 0.82 and 0.78 for the intensity and frequency of precipitation extremes, demonstrating more robust scalings than those identified across climate models. This suggests that the physical processes which constrain the intensification of precipitation extremes to scale with SAT across climate models also hold within individual models across time.

TABLE 2. Assessing improvements in the identification of precipitation-temperature scalings when using LFCA rather than averages over 10, 20 and 30 year periods. The multi-model average increase in the R^2 of the temperature-precipitation scaling relation when using LFCA rather than temporal averages, as well as the percentage of models with a significant (at the 10% level) increase in R^2 based on bootstrapped estimates of the scaling relation uncertainties.

	LFCA vs 10-yr av.	LFCA vs 20-yr av.	LFCA vs 30-yr av.
Rx1 R^2 improvement	0.30	0.14	0.11
% models with sig. improvement	76	42	28
R>99% R^2 improvement	0.38	0.23	0.22
% models with sig. improvement	76	57	38

When compared to using temporal averages over 10, 20 and 30 year periods, we find that first applying LFCA to identify the lowest-frequency changes considerably improves the robustness of the identified scaling relations. Supplementary Figs. S6-7 show the scaling relations identified with LFCA for all ensemble members compared to those in Figs. S8-9 without LFCA. When using temporal averages over longer periods (20, 30 years) fewer observations are available to construct the scaling relations and the R^2 typically increases both with and without LFCA (results using 30 year periods are shown in Supplementary Figs. S10-13). Table 2 documents the improvements of using LFCA, showing that even compared to 30 year averages LFCA improves the R^2 of the identified relations by 0.11 and 0.22 for the intensity and frequency of precipitation extremes

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

respectively. To assess the significance of these improvements in comparison to the uncertainty of the scaling relations, we re-estimate the scaling relations in each ensemble member under 1000 boot-strapped replacements of the different time periods used to estimate the changes. Given this methodological uncertainty, we find that 28% and 38% of climate models exhibit significant improvements in the R^2 of their scaling relation at the 10% level.

d. Discrepancies and constraints on temperature-precipitation scaling rates

Importantly, robustly identifying temperature-precipitation scaling relations within individual ensemble members facilitates a comparison of their rates across climate models (Fig. 5). On average, we find that the rates at which precipitation extremes scale with SAT within individual models is consistent with that identified between models (compare Fig. 5 to Fig. 3), but considerable heterogeneity across models is clear. The coefficient of variation of these rates is 23% and 30% for the intensity and frequency of extremes respectively. Moreover, the inter-model differences in scaling rates are statistically significant at the 5% level for 59 and 65% of model pairs, when assessing uncertainty in the estimation of scaling rates as outlined above. These results demonstrate large biases between models thus diagnosing a source of uncertainty in projections of future precipitation extremes.

To quantify the extent to which these biases contribute to uncertainty in the magnitude of projected precipitation change, we use a simple Monte-Carlo simulation procedure which samples from and combines the different changes in global temperatures and different scaling rates projected by different members of the multi-model ensemble. By assessing the distribution of projected changes when uncertainty is propagated from both sources of errors, or when only one source of error is considered while the other is held fixed at its mean value, we can consider the relative extents to which these two sources contribute to overall uncertainty. Results indicate that under a high-greenhouse forcing scenario, biases in the scaling-rates of the intensity of extremes contribute equally to uncertainty in projected change as do uncertainty in the projected level of warming. For the frequency of extremes, biases in scaling-rates contribute a considerably larger share to the overall uncertainty. For both metrics, biases in the scaling-rates lead to larger right-tailed risks of large increases in precipitation extremes. These results indicate that addressing biases in scaling-

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

rates is at least equally important as constraining the climate sensitivity in reducing uncertainties in projections of future precipitation extremes.

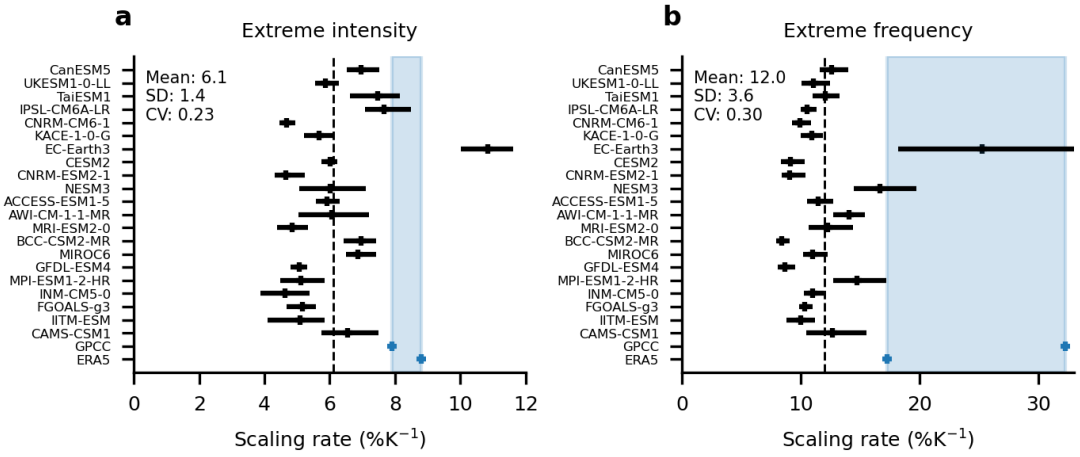

FIG. 5. Comparing scaling rates between precipitation extremes and global temperatures across models and to estimates from observations. (a) Scaling-rates between low-frequency changes in the intensity of daily precipitation extremes estimated for each individual climate model using LFCA, as shown in Fig. 4. 90% confidence intervals on these scaling rates are obtained via boot-strapped replacements of the inter-temporal changes used to estimate the scaling relations, and are shown as horizontal bars. The multi-model mean of the precipitation-temperature scaling rates is shown by the vertical black line with the mean, standard deviation and coefficient of variation across models indicated in the upper left. Estimates of precipitation-temperature scaling rates from two observational products (ERA5 and GPCC) are made by taking the difference between averages over the first and final 20-year period of the observational data period, and are shown in blue. Equivalent data for the frequency of daily precipitation extremes is shown in panel (b). In this case, precipitation percentiles are estimated over the whole historical period for the observational data, and for a similar period (1985-2015) for the climate models (although the scaling rates obtained are largely robust to the period used to estimate percentiles, see Supplementary Fig. S15).

While focused model development and physical theory are crucial avenues to limit these biases, comparison to observations may provide a helpful way to evaluate and select climate models which accurately reproduce observed scaling-rates. We therefore compare the scaling rates identified within individual members of the CMIP-6 ensemble to estimates from observations. Global historical records of daily precipitation are challenging due to large spatial variability and limited

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

ground-based networks, particularly across parts of the developing world, and we therefore use two observational products. We use data from the ERA-5 reanalysis and Global Precipitation Climatology Centre (GPCC) which both focus on daily characteristics. While ERA-5 uses land-station data and satellite observations combined with assimilation techniques from weather forecasting models (Hersbach et al. (2020)), GPCC relies predominantly on land-station data and interpolation (Schneider et al. (2008)). These characteristics enable an assessment of the robustness of observational estimates reflecting uncertainties from both data sources and methodological limitations. Given the relatively shorter time periods available in these observational products (approximately 40 years in each case), we use a simpler methodology to estimate the observed scaling of precipitation extremes with global SAT by taking the difference between averages in precipitation and SAT over the first and final 20 years of available data (note that SAT changes are in both cases derived from ERA5 but adjusted to reflect the time periods of the two precipitation datasets).

This comparison to observations indicates that the historical rates at which the intensification of precipitation extremes have scaled with global SAT exceed those of virtually all the ensemble members in CMIP-6 (observations shown in blue in Fig. 5). The two observational datasets agree closely on the observed scaling rates for the intensity of extremes (approximately 8-9%), in which case only three out of twenty-one models exhibit scaling rates with uncertainty which falls within this range, and only one outlier which actually exceeds this range (EC-Earth3). Regarding the frequency of precipitation extremes, the two observational data-sets exhibit a large range of scaling rates with ERA-5 producing estimates of 17% compared to 32% in GPCC. These discrepancies indicate considerable challenges in the representation of the frequency of precipitation extremes even in historical observational data, a detailed analysis of which is beyond the scope of the present work. However, despite these uncertainties it is still clear that observed rates exceed those identified in CMIP-6. Again, only two models exhibit scaling rates with uncertainties within this range while most models exhibit rates significantly below it. In addition to the assessments of methodological uncertainty already discussed, these results are robust to the exclusion of arid regions with very low initial precipitation values (Supplementary Fig. S14) as well as the use of different base periods to estimate percentiles in the CMIP-6 ensemble (Supplementary Fig. S15).

Combining these observational estimates of temperature-precipitation scaling rates with projected changes in global temperatures from the CMIP-6 ensemble produces an observationally-

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

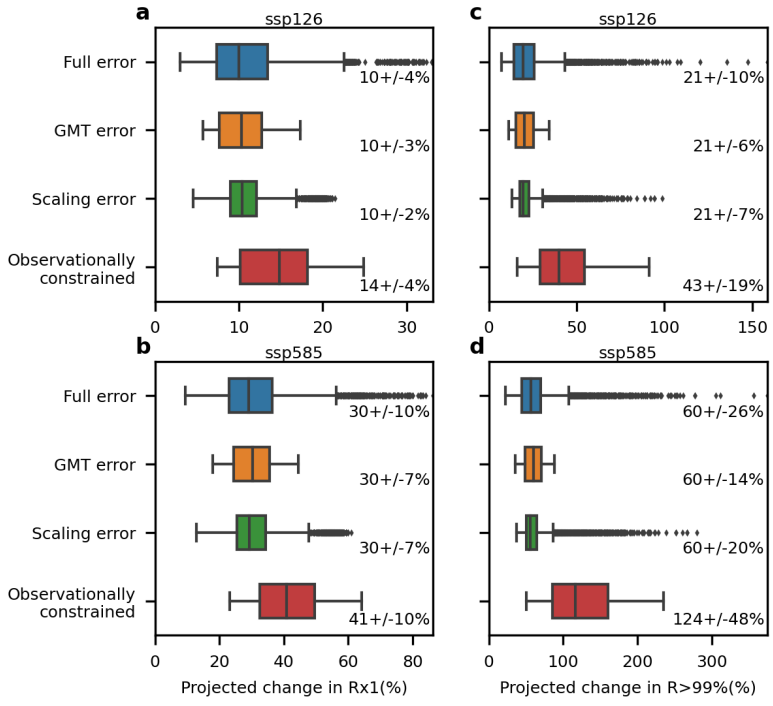

FIG. 6. Quantifying contributions to uncertainty in projected extreme precipitation change, and constraining them with observations. Results show the distribution of projected changes in precipitation extremes when conducting a Monte-Carlo simulation procedure which samples from the range of global SAT change and temperature-precipitation scaling rates projected by each member of the CMIP-6 ensemble. "Full error" indicates simulations where the uncertainty from both sources are propagated, whereas "GMT error" and "Scaling error" indicate cases where uncertainty from only one source was combined with the mean-average of the other component. Finally, the "observationally-constrained" case combines uncertainty from projected SAT change across climate models with the two observational estimates of the temperature-precipitation scaling relations shown in Fig. 5. Results for the intensity of precipitation extremes (RX1) are shown in (a, b) and for the frequency of extremes (R>99%) in (c, d) for both the weaker (SSP126) and stronger (SSP585) greenhouse forcing scenarios. Inset numbers indicate the mean and standard deviation of the distributions.

constrained estimate of the distribution of projected precipitation extremes (Fig. 6). These estimates indicate an upward revision of projected changes in precipitation extremes, by a factor of approximately one third for the intensity of extremes and a doubling for the frequency of extremes. The consistent estimates of scaling rates for the intensity of extremes across the two observational products leads to reductions in the relative error (standard deviation divided by the mean) of the

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

projections, from approximately one third to one quarter in the high forcing scenario. In the case of the frequency of extremes, larger discrepancies between the two observational products mean that relative error in the projections declines by a smaller margin, from 43 to 38%. While a simplified methodology with drawbacks discussed further in the following section, this approach indicates the potential for larger risk from intensifying precipitation extremes than that projected by the CMIP-6 ensemble.

4. Discussion and conclusion

This manuscript uses a pattern-filtering methodology to assess low-frequency changes in the intensity and frequency of daily precipitation extremes in the multi-model ensemble CMIP-6. We find that using LFCA improves agreement on the spatial patterns of projected change. Improvements on using 30-year averages are marginal in the context of strong greenhouse forcing, but under weak greenhouse forcing significant improvements in agreement are found between nearly all model pairs. While this may indicate that LFCA improves the separation of the response of precipitation to human forcing from internal variability, the interpretation of the lowest-frequency change as an anthropogenically-forced signal should be explored further. In particular, SMILEs (Kay et al. (2015)) may provide a fruitful testing ground to evaluate the efficacy of LFCA in this task, and could follow similar studies which demonstrate its strength at detecting forced-responses within individual ensemble members from changes in surface temperature (Wills et al. (2020)). Such tests could encourage the use of LFCA to detect forced changes in historical observations of precipitation where methods capable of detecting responses from single realisations are necessary.

Our work also sheds light on uncertainties in the magnitude of projected changes in precipitation extremes, in particular with regard to how these changes scale with global surface temperatures. Our finding that changes in precipitation extremes scale with global temperatures across members of multi-model ensembles such as CMIP-6 follows a number of other works (Kharin et al. (2013); Li et al. (2021)). However, by using LFCA to robustly identify scaling relations between precipitation extremes and global surface temperature within individual ensemble members, we offer a number of new insights. First, the fact that such scaling relations are identifiable within individual ensemble members across time and scenarios emphasises the robustness of such relations. The fact that on average, they follow the Clausius-Clapeyron relation (for the intensity of extremes) emphasises

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

the fact that thermodynamic drivers of precipitation extremes are dominant determinants of their global change.

Second, our approach reveals strong heterogeneity between different climate models in the rates at which precipitation extremes scale with global temperatures. This diagnoses a source of the uncertainty in projections of future precipitation extremes which we estimate to be equally large as contributions from uncertainty in the climate sensitivity. Our analysis using Monte-Carlo simulations to combine global surface temperature changes and global precipitation scaling rates is naturally simplified and does not explicitly consider the driving physical processes. Nevertheless it is justified by the fact that temperature-precipitation scaling rates appear largely independent of SAT change (Fig. A2), and its findings demonstrate that constraining these scaling rates is equally important as constraining climate sensitivity for limiting uncertainties in future projections. Large efforts have been placed in reducing uncertainties in climate sensitivity (Sherwood et al. (2020)). Given the importance of precipitation extremes for impacts, our results may encourage investing similar efforts using a combination of physical theory, observations and potentially even paleo-climatic evidence (Schmidt et al. (2013); Carmichael et al. (2018)) to constrain the sensitivity of secondary climatic characteristics like precipitation extremes to global SAT change.

Third, we find that the rates at which precipitation extremes are projected to scale with global temperatures under-estimate those observed historically in nearly all climate models. This is a concerning finding, indicating that there is risk of larger increases in precipitation extremes than those currently projected by models, particularly in terms of their frequency. These results are qualitatively consistent with earlier findings that the observed intensification of precipitation extremes were under-estimated by older generations of climate models (Liu et al. (2009); Shiu et al. (2012)), although different methods for measuring extremes make our findings difficult to compare quantitatively. While the short observational record may contribute to errors in the quantification of historical scaling-rates, the fact that climate models indicate a robust-scaling of changes across time and scenarios may add confidence to this assessment. Our combination of observed scaling-rates of precipitation extremes with projected changes in SAT allow an observationally-constrained estimate of future changes in extremes which reduces the uncertainty and increases the magnitude of projected change. Constraining projections with observations is a rapidly growing field, showing recent promise with precipitation (O'Gorman (2012); Zhang and Soden (2019);

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Shiogama et al. (2022); Chen et al. (2022); Zhang et al. (2022); Thackeray et al. (2022); Schewe and Levermann (2022)). In particular, our methodology using temperature-precipitation scaling rates sheds complementary light on a recent method which uses observed historical changes to constrain future projections (Thackeray et al. (2022)). While the method of Thackeray et al. (2022) reduces uncertainty in projected precipitation extremes, it does not find observed changes to be larger than those projected by models, and the observational-constraint therefore does not increase the projected magnitude of change. This discrepancy to the results of our analysis and previous studies (Liu et al. (2009); Shiu et al. (2012)) indicates that separating the different sources of uncertainty in projected precipitation change, namely uncertainty in projected warming as well as the rates at which precipitation extremes scale with global temperatures, is important not only for our qualitative understanding but also for the quantitative results of constraining projections with observations. Nevertheless, our analysis constraining projections with observations is primarily for conceptual purposes, and we recognise that combining our results with a constraint on the climate sensitivity would be necessary to arrive at a fully-constrained projection of precipitation extremes as that provided by (Thackeray et al. (2022)). Indeed, discrepancies between observed and projected historical warming likely underlie the differences in projected precipitation when using these two methods to constrain them. Further analysis in this regard is an interesting avenue for future research, but beyond the scope of the present manuscript.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Acknowledgments. MK and LW received funding from the Volkswagen foundation. SL received funding from the German Research Foundation (DFG, project number 427397136) and from the German Federal Ministry of Education and Research (BMBF, project number 01LP1907A). AL received funding from the Horizon 2020 Framework Programme of the European Union (grant agreement number 820712).

Data availability statement. Raw CMIP6 data is available from https://esgf-node.llnl.gov/projects/cmip6/. Code for low-frequency component analysis is available from https://github.com/rcjwills/lfca. All other data and code is available form the authors upon request.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0492.1.

APPENDIX A

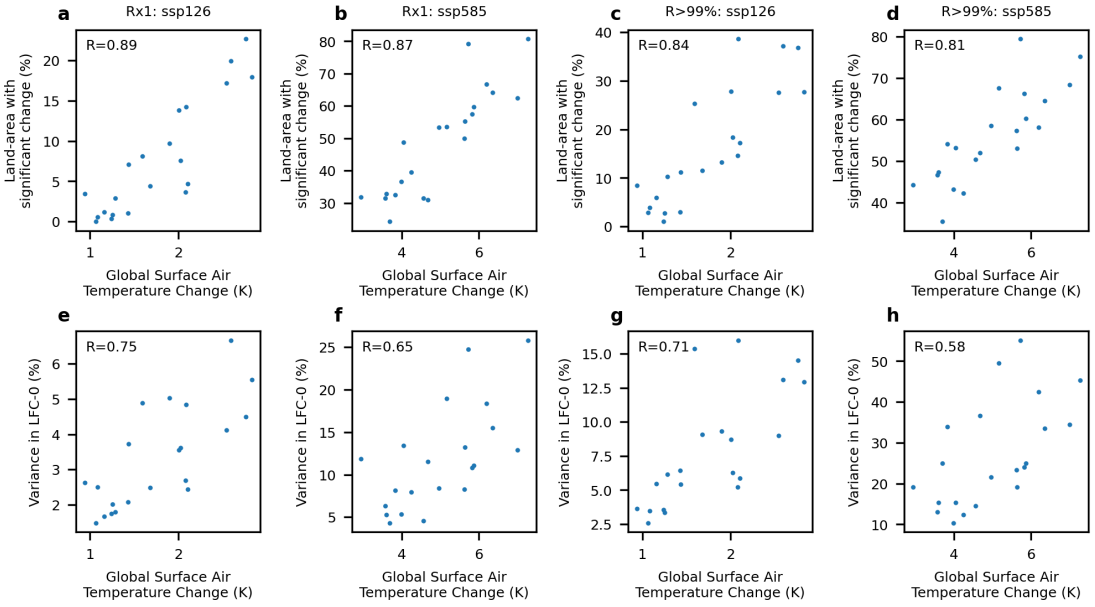

FIG. A1. Scatter plots and pearson correlation values (R) for the fraction of land-area with significant changes (a-d) and the percentage of spatio-temporal variance explained by the lowest-frequency component (e-h) against changes in global surface air temperatures. Results are shown for the intensity (Rx1) and frequency (R>99%) of daily precipitation extremes, and for the low (ssp126) and high (ssp585) greenhouse forcing scenarios.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

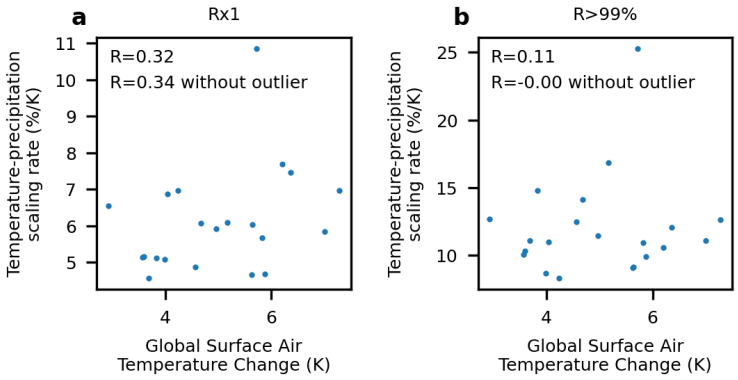

FIG. A2. Scatter plots and pearson correlation values between the temperature-precipitation scaling rates and global SAT change in the high-forcing scenario (ssp585) identified in CMIP-6 climate models.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

References

Aalbers, E. E., G. Lenderink, E. van Meijgaard, and B. J. van den Hurk, 2018: Local-scale changes in mean and heavy precipitation in western europe, climate change or internal variability? *Climate Dynamics*, **50**, 4745–4766.

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419 (6903)**, 228–232.

Blanusa, M. L., C. J. López-Zurita, and S. Rasp, 2023: Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes. *Climate Dynamics*, 1–15.

Boulange, J., N. Hanasaki, D. Yamazaki, and Y. Pokhrel, 2021: Role of dams in reducing global flood exposure under climate change. *Nature Communications*, **12 (1)**, 417, https://doi.org/10.1038/s41467-020-20704-0, URL https://doi.org/10.1038/s41467-020-20704-0.

Carmichael, M. J., R. D. Pancost, and D. J. Lunt, 2018: Changes in the occurrence of extreme precipitation events at the paleocene–eocene thermal maximum. *Earth and Planetary Science Letters*, **501**, 24–36.

Chen, H., and J. Sun, 2017: Contribution of human influence to increased daily precipitation extremes over china. *Geophysical Research Letters*, **44 (5)**, 2436–2444, https://doi.org/https://doi.org/10.1002/2016GL072439, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072439, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL072439.

Chen, Z., T. Zhou, X. Chen, W. Zhang, L. Zhang, M. Wu, and L. Zou, 2022: Observationally constrained projection of afro-asian monsoon precipitation. *Nature Communications*, **13 (1)**, 2552.

Davenport, F. V., M. Burke, and N. S. Diffenbaugh, 2021: Contribution of historical precipitation change to us flood damages. *Proceedings of the National Academy of Sciences*, **118 (4)**, https://doi.org/10.1073/pnas.2017524118, URL https://www.pnas.org/content/118/4/e2017524118, https://www.pnas.org/content/118/4/e2017524118.full.pdf.

Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting north american climate over the next 50 years: Uncertainty due to internal variability. *Journal of Climate*, **27 (6)**,

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

2271 – 2296, https://doi.org/10.1175/JCLI-D-13-00451.1, URL https://journals.ametsoc.org/view/journals/clim/27/6/jcli-d-13-00451.1.xml.

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. *Geoscientific Model Development*, **9 (5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, URL https://gmd.copernicus.org/articles/9/1937/2016/.

Fischer, E. M., and R. Knutti, 2016: Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, **6 (11)**, 986–991, https://doi.org/10.1038/nclimate3110, URL https://doi.org/10.1038/nclimate3110.

Fischer, E. M., J. Sedláček, E. Hawkins, and R. Knutti, 2014: Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, **41 (23)**, 8554–8562.

Frame, D. J., S. M. Rosier, I. Noy, L. J. Harrington, T. Carey-Smith, S. N. Sparrow, D. A. Stone, and S. M. Dean, 2020: Climate change attribution and the economic costs of extreme weather events: a study on damages from extreme rainfall and drought. *Climatic Change*, **162**, 781–797.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049.

Hsiang, S. M., M. Burke, and E. Miguel, 2013: Quantifying the influence of climate on human conflict. *Science*, **341 (6151)**, 1235 367, https://doi.org/10.1126/science.1235367.

Kay, J. E., and Coauthors, 2015: The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, **96 (8)**, 1333 – 1349, https://doi.org/10.1175/BAMS-D-13-00255.1, URL https://journals.ametsoc.org/view/journals/bams/96/8/bams-d-13-00255.1.xml.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the cmip5 ensemble. *Climatic change*, **119**, 345–357.

King, A. D., and Coauthors, 2015: The timing of anthropogenic emergence in simulated climate extremes. *Environmental Research Letters*, **10 (9)**, 094 015.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Kirchmeier-Young, M. C., and X. Zhang, 2020: Human influence has intensified extreme precipitation in north america. *Proceedings of the National Academy of Sciences*, **117 (24)**, 13 308–13 313, https://doi.org/10.1073/pnas.1921628117, URL https://www.pnas.org/content/117/24/13308, https://www.pnas.org/content/117/24/13308.full.pdf.

Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new cmip5 climate model projections. *Nature climate change*, **3 (4)**, 369–373.

Kotz, M., A. Levermann, and L. Wenz, 2022: The effect of rainfall changes on economic production. *Nature*, https://doi.org/10.1038/s41586-021-04283-8.

Kotz, M., L. Wenz, and A. Levermann, 2021: Footprint of greenhouse forcing in daily temperature variability. *Proceedings of the National Academy of Sciences*, **118 (32)**, https://doi.org/10.1073/pnas.2103294118, URL https://www.pnas.org/content/118/32/e2103294118, https://www.pnas.org/content/118/32/e2103294118.full.pdf.

Lange, S., and Coauthors, 2020: Projecting exposure to extreme climate impact events across six event categories and three spatial scales. *Earth's Future*, **8 (12)**, e2020EF001 616, https://doi.org/https://doi.org/10.1029/2020EF001616, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001616, e2020EF001616 10.1029/2020EF001616, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001616.

Li, C., F. Zwiers, X. Zhang, G. Li, Y. Sun, and M. Wehner, 2021: Changes in annual extremes of daily temperature and precipitation in cmip6 models. *Journal of Climate*, **34 (9)**, 3441–3460.

Liang, X.-Z., and Coauthors, 2017: Determining climate effects on us total agricultural productivity. *Proceedings of the National Academy of Sciences*, **114 (12)**, E2285–E2292, https://doi.org/10.1073/pnas.1615922114, URL https://www.pnas.org/content/114/12/E2285, https://www.pnas.org/content/114/12/E2285.full.pdf.

Liu, S. C., C. Fu, C.-J. Shiu, J.-P. Chen, and F. Wu, 2009: Temperature dependence of global precipitation extremes. *Geophysical Research Letters*, **36 (17)**.

Madakumbura, G., C. C.W. Thackeray, and J. J. Norris, 2021: Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nature Communications*, **12 (3944)**, https://doi.org/10.1073/pnas.1921628117.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Masson-Delmotte, V., and Coauthors, 2021: IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press. In Press.*

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470 (7334)**, 378–381, https://doi.org/10.1038/nature09763, URL https://doi.org/10.1038/nature09763.

O'Gorman, P. A., 2012: Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, **5 (10)**, 697–700, https://doi.org/10.1038/ngeo1568, URL https://doi.org/10.1038/ngeo1568.

O'Gorman, P. A., and T. Schneider, 2009: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, **106 (35)**, 14 773–14 777.

Pfahl, S., P. A. O'Gorman, and E. M. Fischer, 2017: Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, **7 (6)**, 423–427, https://doi.org/10.1038/nclimate3287, URL https://doi.org/10.1038/nclimate3287.

Santer, B. D., K. E. Taylor, T. M. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dynamics*, **12**, 77–100.

Santer, B. D., and Coauthors, 2007: Identification of human-induced changes in atmospheric moisture content. *Proceedings of the National Academy of Sciences*, **104 (39)**, 15 248–15 253.

Schewe, J., and A. Levermann, 2022: Sahel rainfall projections constrained by past sensitivity to global warming. *Geophysical Research Letters*, e2022GL098286.

Schmidt, G., and Coauthors, 2013: Using paleo-climate comparisons to constrain future projections in cmip5. *Climate of the Past Discussions*, **9 (1)**.

Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf, 2008: Global precipitation analysis products of the gpcc. *Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation*, **112**.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Schumacher, R. S., 2017: Heavy rainfall and flash flooding. *Oxford research encyclopedia of natural hazard science*.

Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, **7 (10)**, 703–708, https://doi.org/10.1038/ngeo2253, URL https://doi.org/10.1038/ngeo2253.

Sherwood, S. C., and Coauthors, 2020: An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, **58 (4)**, e2019RG000 678.

Shiogama, H., M. Watanabe, H. Kim, and N. Hirota, 2022: Emergent constraints on future precipitation changes. *Nature*, **602 (7898)**, 612–616.

Shiu, C.-J., S. C. Liu, C. Fu, A. Dai, and Y. Sun, 2012: How much do precipitation extremes change in a warming climate? *Geophysical Research Letters*, **39 (17)**.

Thackeray, C. W., A. Hall, J. Norris, and D. Chen, 2022: Constraining the increased frequency of global precipitation extremes under warming. *Nature Climate Change*, **12 (5)**, 441–448.

Thiery, W., and Coauthors, 2021: Intergenerational inequities in exposure to climate extremes. *Science*, **374 (6564)**, 158–160, https://doi.org/10.1126/science.abi7339.

Thomas, B. F., and J. S. Famiglietti, 2019: Identifying climate-induced groundwater depletion in grace observations. *Scientific reports*, **9 (1)**, 4124.

von Uexkull, N., M. Croicu, H. Fjelde, and H. Buhaug, 2016: Civil conflict sensitivity to growing-season drought. *Proceedings of the National Academy of Sciences*, **113 (44)**, 12 391–12 396, https://doi.org/10.1073/pnas.1607542113, URL https://www.pnas.org/content/113/44/12391, https://www.pnas.org/content/113/44/12391.full.pdf.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe, 2014: The inter-sectoral impact model intercomparison project (isi–mip): Project framework. *Proceedings of the National Academy of Sciences*, **111 (9)**, 3228–3232, https://doi.org/10.1073/pnas.1312330110, URL https://www.pnas.org/content/111/9/3228, https://www.pnas.org/content/111/9/3228.full.pdf.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.

Willner, S. N., A. Levermann, F. Zhao, and K. Frieler, 2018: Adaptation required to preserve future high-end river flood risk at present levels. *Science Advances*, **4 (1)**, eaao1914, https://doi.org/10.1126/sciadv.aao1914.

Wills, R. C., T. Schneider, J. M. Wallace, D. S. Battisti, and D. L. Hartmann, 2018: Disentangling global warming, multidecadal variability, and el niño in pacific temperatures. *Geophysical Research Letters*, **45 (5)**, 2487–2496, https://doi.org/https://doi.org/10.1002/2017GL076327, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076327, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL076327.

Wills, R. C. J., D. S. Battisti, K. C. Armour, T. Schneider, and C. Deser, 2020: Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *Journal of Climate*, **33 (20)**, 8693 – 8719, https://doi.org/10.1175/JCLI-D-19-0855.1, URL https://journals.ametsoc.org/view/journals/clim/33/20/jcliD190855.xml.

Wood, R., and R. Ludwig, 2020: Analyzing internal variability and forced response of subdaily and daily extreme precipitation over europe. *Geophysical Research Letters*, **47 (17)**, e2020GL089 300.

Zhang, B., and B. J. Soden, 2019: Constraining climate model projections of regional precipitation change. *Geophysical Research Letters*, **46 (17-18)**, 10 522–10 531.

Zhang, W., K. Furtado, T. Zhou, P. Wu, and X. Chen, 2022: Constraining extreme precipitation projections using past precipitation variability. *Nature Communications*, **13 (1)**, 6319.

Zhang, X., H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min, 2013: Attributing intensification of precipitation extremes to human influence. *Geophysical Research Letters*, **40 (19)**, 5252–5257, https://doi.org/https://doi.org/10.1002/grl.51010, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.51010, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/grl.51010.

Ziegler, A. D., J. Sheffield, E. P. Maurer, B. Nijssen, E. F. Wood, and D. P. Lettenmaier, 2003: Detection of intensification in global-and continental-scale hydrological cycles: Temporal scale of evaluation. *Journal of Climate*, **16 (3)**, 535–547.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0492.1.