Potential to Constrain Projections of Hot Temperature Extremes[✉]

ALEKSANDRA BORODINA, ERICH M FISCHER, AND RETO KNUTTI

Institute for Atmosphere and Climate Science, ETH Zurich, Zurich, Switzerland

(Manuscript received 25 November 2016, in final form 16 August 2017)

ABSTRACT

Projected changes in temperature extremes, such as regional changes in the intensity and frequency of hot extremes, differ strongly across climate models. This study shows that this disagreement can be partly explained by discrepancies in the representation of the present-day temperature distribution, motivating the evaluation of models with observations. By evaluating climate models on carefully selected metrics, the models that are more likely to be reliable for long-term projections of temperature extremes are identified. The study found that frequencies of hot extremes are likely to increase at a higher rate than the multimodel mean estimate over large parts of the Northern Hemisphere and Australia. This implies that a higher degree of adaptation is required for a given global temperature target. It also found that projected changes in the intensity of hot extremes can be constrained in several regions, including Australia, central North America, and north Asia. In many other regions, large internal variability can often hamper model evaluation. For both aspects—the intensity and the frequency of hot extremes—the total area over which the constraints can be implemented is limited by the quality and completeness of observations. Thereby, this study highlights the importance of long-term, high-quality, and easily accessible observational records for model evaluation, which are vital to ultimately reduce uncertainties in projections of temperature extremes.

1. Introduction

Hot temperature extremes are projected to intensify and become more frequent in the future (e.g., Kharin et al. 2007, 2013; Meehl and Tebaldi 2004; Alexander and Tebaldi 2012; Sillmann et al. 2013b; Perkins et al. 2012; Cowan et al. 2014; Coumou and Robinson 2013; Coumou et al. 2013; Fischer et al. 2011; Alexander and Arblaster 2009). However, for a given scenario climate models disagree on the magnitude of the future change in intensity and frequency because of model deficiencies (Boberg and Christensen 2012; Christensen and Boberg 2012; Cattiaux et al. 2013) as well as internal variability. The large disagreement across climate models presents a key challenge for the assessment of climate change impacts and hence the development of adaptation and mitigation strategies. The observational data provide an opportunity to reduce the disagreement across models by means of so-called *observational* or *emergent* constraints

(e.g., Fasullo and Trenberth 2012; Knutti et al. 2006; Knutti and Tomassini 2008; Hall and Qu 2006; Holland and Bitz 2003; Boé et al. 2009; Son et al. 2010; Schaller et al. 2011; Mahlstein and Knutti 2012; O’Gorman 2012; Massonnet et al. 2012; Cox et al. 2013; Wenzel et al. 2016; Loeb et al. 2016; Stegehuis et al. 2013; Tian 2015; Su et al. 2014; Huber et al. 2011).

Observational constraints aim at identifying discrepancies in simulations of the present-day climate that lead to large disagreement across model projections. In multimodel ensembles, observational constraints emerge as statistical relationships between metrics derived from simulations of the past or present day and projections of interest, as illustrated in Fig. 1. To be able to use this relationship, there must be an understanding of why such a relationship exists, as correlations may arise simply by chance given the large output provided by climate models (Huber et al. 2011; Caldwell et al. 2014). Then, the observations of the present day (green line in Fig. 1) are used to evaluate models and to produce a recalibrated ensemble of models (red range b’) with significantly smaller spread across ensemble members than across the full range of models available (blue range a’). By doing so, we are making an implicit assumption that the disagreement in present-day simulations and the variations

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-16-0848.s1>.

Corresponding author: aleksandra.borodina@env.ethz.ch

FIG. 1. Schematic diagram illustrating an observational constraint. Full multimodel spread and the selected models for the present-day metric (a and b) and projections (a' and b'). Models are selected based on their agreement in the present-day metric with the observational estimate (green line).

in model projections, in part, stem from differences in the representation of processes in models. By applying observational constraints, we aim to eliminate models that poorly capture these present-day local processes, as they are assumed to be inadequate for projections.

In this study, we identify and implement observational constraints on projections of intensity and frequency of hot temperature extremes with the aim to provide ensembles of models with better agreement across its members. In a warming world, intensification of hot extremes often implies that hot events of certain intensity will become more frequent. However, projections of these two aspects (intensity as opposed to frequency) are considered separately, because different biases in the representation of the present-day climate are more relevant for projections of one than the other. The challenge is to identify regions where the identified biases are the main contributor to the disagreement in model projections. Evaluation of models with observations in those regions would allow for identifying a more likely range of projections with improved agreement across its members.

Temperature extremes were shown to intensify at a rate approximately following changes in the local mean temperatures in many parts of the world (e.g., Kharin et al. 2013). However, this is not a good approximation for the regions where the scale and shape of the temperature distribution are expected to change as a consequence of climate change (e.g., Kharin and Zwiers 2000, 2005). Katz and Brown (1992) were first to suggest that changes in extremes can be quite different from what is expected from a uniform shift of the climatology. Since then, more studies have demonstrated the

asymmetry in warming between the tails of the distribution and the mean in observations (e.g., Klein Tank and Können 2003; Christidis et al. 2005; Robeson 2004) and model simulations (e.g., Schär et al. 2004; Fischer and Schär 2009; Holmes et al. 2016). The local asymmetry in warming is represented by the model's ability to capture relevant processes and feedbacks. For example, in regions where strong drying occurs (e.g., central Europe), the hot tail of the temperature distribution warms faster than the local mean temperature, leading to a widening of the temperature distribution. This has been largely attributed to reduced evaporative cooling from the land surface caused by soil moisture deficit (Gregory and Mitchell 1995; Schär et al. 2004; Alexander and Perkins 2013; Seneviratne et al. 2006; Fischer et al. 2012). Hence, to study the future relative intensity of hot extremes, we focus on the scaling (the regression) of the annual maximum temperatures (TXx) with local mean summer temperatures (TXx scaling). High-magnitude TXx scalings occur in cases where the rate of warming of the tail (TXx) is very different from the rate of warming in the center of the distribution (local mean warming).

The changing frequency (as opposed to intensity) of hot extremes is another key aspect of the future climate. Frequency is often defined relative to a local percentile threshold of the present-day climatology. A uniform shift of the temperature distribution toward warmer temperatures will lead to an increased frequency of events exceeding such a fixed threshold, and therefore it can account for most of the projected increase in frequencies (Coumou and Robinson 2013; Coumou et al. 2013; Fischer and Knutti 2015; Griffiths et al. 2005; Simolo et al. 2011; Ballester et al. 2010). Consequently, the disagreement across models on the magnitude of future frequencies can be partly attributed to the model-specific shape of the underlying temperature distribution (e.g., Simolo et al. 2011; Lovejoy 2014) and its evolution under anthropogenic forcing (i.e., asymmetric changes in the tails as discussed earlier) (e.g., Della-Marta et al. 2007; Ballester et al. 2010). However, their relative importance depends on the region. Hence, we search for regions where the present-day distribution is the main contributor to the disagreement in model-simulated frequencies. Then we evaluate models on their ability to accurately simulate the shape of the present-day temperature distribution in these regions.

This study provides a global assessment on where the uncertainties in projections of intensities and frequencies can be narrowed with observational constraints. This paper consists of two main parts. In the first part we focus on projections of intensity of hot extremes (by examining the TXx scaling) and in the second part we

focus on the frequency of hot extremes (by examining the width of the tail of the present-day distribution). We explore the main challenges that prevent model evaluation, as these are important to consider when applying this method to other variables. Improved model agreement for both, future intensity and frequency, will provide a more complete picture of the future climate. This improvement is needed for more rigorous impact assessments (Murray and Ebi 2012).

2. Data and methods

a. Datasets

We use model-simulated summer daily maximum temperature (TX) from phase 5 of the Coupled Model Intercomparison Project (CMIP5) in addition to the 10-member CESM initial condition ensemble produced at ETH Zurich (Fischer et al. 2013). A total of 28 models that provide all the necessary data for the historical and representative concentration pathway 8.5 (RCP8.5) scenario simulations are analyzed (Table S1 in the online supplemental material). Model simulations are evaluated against ERA-Interim (1979–2014) (Dee et al. 2011), the NCEP Twentieth Century Reanalysis (1951–2011) (Compo et al. 2011), and the NCEP–DOE AMIP-II reanalysis (1979–2014) (Kanamitsu et al. 2002). In the reanalyses, TX is approximated by the diurnal maximum of the shortest time step available. Observational datasets used in the analysis are the Hadley Centre Global Climate Extremes Index 2 (HadEX2) (1951–2010) (Donat et al. 2013a), the Global Historical Climatology Network–Daily (GHCND)-based climate extremes indices (GHCNDEX) (1951–2014) (Donat et al. 2013b), and the Hadley Centre–NCDC-developed GHCND (HadGHCND) (1949–2011) (Caesar et al. 2006). Seasonal mean temperatures were estimated from the Climatic Research Unit air temperature (CRUTEM), version 4.3.0, dataset for grid points where at least June and August (December and February for the Southern Hemisphere) were available (Jones and Moberg 2003). Our analysis focuses on land areas (grid boxes with at least 50% land) computed from the land–ocean mask of the CESM.

To constrain the projected intensification of TXx, we are interested in the rate at which TXx is increasing relative to the mean temperatures. TXx values in HadEX2 and GHCNDEX are readily available on a $2.5^\circ \times 3.75^\circ$ grid and represent interpolated station extremes. In models and reanalyses TXx values are calculated on their native grids (Table S3) and then regridded onto a $2.5^\circ \times 3.75^\circ$ grid using a first-order conservative remapping procedure (Jones 1999). This order ensures a fairer comparison of models with observational datasets such as HadEX2 and

GHCNDEX that interpolate station extremes onto a coarse grid. It is important to note that the HadGHCND dataset first interpolates daily data onto a $2.5^\circ \times 3.75^\circ$ grid and then computes TXx at a gridpoint scale. For each observational dataset, grid boxes are masked where more than 20% of the TXx time series is missing. When we evaluate models against all three observational datasets (HadEX2, GHCNDEX, and HadGHCND), we analyze grid boxes where all three observational datasets have coverage.

To constrain the projected frequencies, the width of the distribution is key. Hence, to evaluate the distribution of daily maximum temperatures (TX distribution), we use the 95th-to-75th-percentile difference. The order of operation needed to evaluate the frequencies of hot extremes is different from that for intensities of hot extremes discussed earlier. Since raw daily data are required, the analysis is restricted to HadGHCND, as it is the only gridded global dataset with raw temperature data at daily resolution. To ensure a direct comparison between models and HadGHCND is possible, the daily maximum temperatures are first regridded onto the $2.5^\circ \times 3.75^\circ$ grid using the same remapping procedure and then the rest of the calculations are performed on a common grid (e.g., percentiles). Grid boxes with less than 95% of all time steps in summer over the period 1979–2010 ($n \approx 2900$) are masked.

We focus on 3°C global annual mean warming in each model (5-yr running mean relative to the period 1979–2010). Models with at least four initial condition members are used to estimate the internal variability of the quantities at hand (e.g., TXx scaling). This is done by computing the multimodel mean of the standard deviations calculated only from models with multiple runs.

b. Methods for TXx scaling estimates

Gridpoint TXx scaling estimates, the regression slope of TXx against local summer mean temperatures, are calculated using a Theil–Sen slope estimator because it is more robust than simple linear ordinary least squares regression in the presence of outliers (von Storch and Zwiers 1999). We then calculate the area median TXx scaling, which is robust in the presence of outliers, over the Giorgi regions (Giorgi and Francisco 2000) and some larger regions (Table S2). For all of the regions, TXx is regressed against June–August (JJA) mean temperature means except for southern South America (SSA), southern Africa, and Australia, where December–February (DJF) mean temperatures are used.

The historical TXx scaling estimates in models are calculated for the period 1951–2014 (and 1901–2014 for Fig. 4). For models with different realizations, we calculate the TXx scalings for each individual realization

and average them prior to the selection of models. The temporal coverage in observational datasets varies from dataset to dataset (periods provided by individual datasets are in [section 2a](#)). We therefore use all the years with an existing record between 1951 and 2014 provided by an individual dataset to estimate the historical TXx scaling. When exploring the potential for constraining at different warming levels (as for [Fig. 3](#)), we replace time by global mean warming (5-yr running mean relative to 1979–2010).

Individual models are ranked according to their agreement in the historical TXx scaling with three datasets (HadEX2, HadGHCND, and GHCNDEX). For each of the regions, the agreement is quantified as a squared bias between the observed and simulated area median TXx scaling. Biases are calculated relative to individual observational datasets and then averaged to select the best-performing models relative to all three datasets used. From here on, the “constrained” projection (b' in [Fig. 1](#)) is the projection resulting from the selected seven models only. The number of models selected is a compromise between the total number of models available and the necessary number of models needed to sample the uncertainty. Too many models will weaken the constraint, whereas too few (equivalent to too aggressive weighting) are likely to lead to overconfident results that are not robust, particularly if the number of models is small ([Knutti et al. 2017](#)). The number of models selected does somewhat affect the spread of the constrained projection; however, it does not undermine the general conclusions that the constrained projections are on the higher/lower end of the full ensemble ([Borodina et al. 2017](#)).

c. Method for threshold exceedances

We quantify the annual frequency of exceeding the 95th percentile of the daily summer temperature. We use absolute temperatures to calculate the percentiles at individual grid points for each model and observational dataset individually. The percentile threshold is computed over the entire length of the reference period 1979–2010. Then, the frequency within each model is computed by counting the total number of days exceeding the threshold for each year of the simulation.

The projected simulated frequencies are then computed as a mean of the 32-yr period centered around the 5-yr period with 3°C global mean warming. To disentangle the uncertainties in the projected frequencies of hot extremes, we compare model-simulated frequencies with frequencies resulting from simply shifting the 32-yr reference period (1979–2010) by a fixed warming pattern. The first estimate is calculated from shifting the reference period by a warming pattern consistent with

3°C multimodel mean warming (multimodel shift). The second estimate is calculated by shifting the reference period by a warming pattern consistent with 3°C warming in the respective individual model (single-model shift). The frequencies estimated by shifting the reference period are then computed by taking the mean of the 32-yr period.

We calculate the area-weighted mean of frequencies (simulated or estimated) for the regions of interest (Table S2). Summer JJA temperatures are used for the Northern Hemisphere and DJF temperatures are used for the Southern Hemisphere. The metric used to evaluate the present-day distribution in models relative to observations is the temperature difference between the daily 95th and 75th summer temperature percentiles computed from the 1979–2010 reference period, that is, the 95-to-75th-percentile difference. The metric is calculated at each grid point and then averaged across the respective region for each model individually.

3. Constraints on intensity of hot extremes

a. Scaling of hot extremes with local warming

We find that the historical TXx scaling ([Fig. 2a](#)) estimated from 1951 to 2014 is often a good approximation for the long-term TXx scaling (1951 until 3°C warming) ([Fig. 2b](#)). In preindustrial control simulations, TXx values scale with mean summer temperatures because extremely hot days are more likely to occur during particularly hot summers rather than cold summers (not shown). This scaling becomes much clearer in forced simulations (such as RCP8.5) than in preindustrial control simulations, since the warming signal induces increasing TXx values and mean summer temperatures. Model differences in TXx scaling are illustrated for a grid point in central North America ([Fig. 2a](#)), where one model (cyan, eight model realizations) is showing substantially higher long-term TXx scaling (1951–2100) than another model (red, five model realizations). The long-term TXx scaling is well defined as several realizations of the same model falling on top of each other ([Fig. 2a](#)). However, the historical TXx scaling is more affected by large internal variability caused by the short period and small warming signal. This makes an accurate estimate of the historical TXx scaling challenging. For the grid point considered in [Fig. 2a](#), the relationship between TXx and mean temperature is linear. Several studies found that regionally aggregated TXx values scale linearly with global mean temperatures in many of the regions analyzed in this work ([Vogel et al. 2017](#); [Seneviratne et al. 2016](#)), suggesting that the assumption of linearity is reasonable.

Models with low historical median TXx scaling within the region of central North America (x axis in [Fig. 2b](#))

FIG. 2. (a) Regression between TXx values and summer mean temperatures at a grid point in CNA under the RCP8.5 scenario for 1951–2100. One model represents eight initial condition members (cyan) and another five initial condition members (red). The one-to-one slope is indicated (black dashed line). (b) Relationship between

also tend to have a low long-term TXx scaling (y axis) for the region of central North America. This relationship between historical and long-term TXx scaling can be used for an observational constraint of regional TXx projections. Models that are in better agreement with observations are retained. Models that fail to accurately capture the TXx scaling as represented in observational datasets are considered inadequate for projections of TXx scaling and are therefore excluded from the constrained ensemble.

b. Criteria for constraining

Three factors could undermine the use of observational constraints: a weak relationship between historical and long-term responses, large internal variability preventing an accurate historical TXx scaling estimate, and limitations of the observational datasets. We formulate three criteria to address these issues. The first criterion (criterion 1) ensures a strong relationship between the historical and long-term TXx scaling across models. For the region of central North America (CNA), this relationship is strong ($R = 0.83$). However, this is not always the case for other regions. We average multiple initial condition runs of the same model prior to the computation of correlation coefficients across models. To apply a constraint, we specify the criterion that the correlation needs to be higher than 0.6. We expect the relationship to become clearer with increasing length of available data and a stronger warming signal.

We do not expect realizations of the same model to have identical TXx scaling estimates over a short period of a few decades as a result of internal variability (Perkins and Fischer 2013; Perkins-Kirkpatrick et al. 2017). Hence, we estimate the uncertainty range of TXx scaling estimates as a result of internal variability by analyzing models with multiple initial condition members (the mean of standard deviations across all models

the historical TXx scaling (1951–2014) and TXx scaling at 3°C global mean warming in CNA, aggregated as the median of TXx scaling across all grid points. Each dot is one model simulation. Models that have multiple simulations are color coded. Solid vertical lines are observational estimates from HadEX2 (magenta), HadGHCND (red), and GHCNDEX (yellow). Dashed vertical lines are estimates from ERA-Interim (green), NCEP-2 (blue), and the NCEP Twentieth Century Reanalysis (black). (c) Constraining criteria for the CNA region. Criterion 1 (red): correlation across models between TXx scaling as a function of global mean warming and the long-term TXx scaling at 3°C. Criterion 2 (blue): ratio between the natural variability of TXx scaling estimates and the full range simulated by models. Vertical lines show global mean warming estimates from reanalyses.

that have multiple members). Better sampling and a larger warming signal reduce the importance of internal variability (Fischer and Knutti 2014). Thus, we estimate the warming level needed to be able to derive a robust estimate of a model's historical TXx scaling by extending the historical to a certain future warming level rather than 2014. Figure 2c illustrates that with increasing warming levels, the correlation between the extended historical and long-term TXx scaling increases (red dots) and the fraction of variability-induced uncertainty in TXx scaling decreases (blue dots). To this end we introduce a second criterion (criterion 2), ensuring that the fractional uncertainty of the TXx scaling estimate induced by internal variability is less than a third of the total spread of the TXx scaling across models (below 0.33 on the right axis in Fig. 2c). Because the spread caused by internal variability is irreducible, this threshold implies that the model spread can be reduced to a third of the full ensemble width if a robust observational estimate exists. In principle, both thresholds used for criteria 1 and 2 can be relaxed or made stricter (i.e., allowing lower/higher correlations or larger/lower internal variability). Relaxing these criteria would allow more regions to be constrained at the cost of a smaller reduction in the model spread in these regions. Therefore, depending on the application, the thresholds should be chosen to ensure that the reduction in spread is substantial.

The two criteria mentioned above are necessary but are not sufficient conditions to narrow down the disagreement across models by means of observational constraints. Observational uncertainty exists because of potential sampling errors, inhomogeneities, and gridding procedures (Alexander et al. 2006; Donat et al. 2013a; Alexander 2016). It needs to be accounted for (at least partly by considering several datasets) and has to be small enough to justify the model selection. We find that reanalyses agree well with observational datasets in historical TXx scaling (the discrepancy between datasets is smaller than uncertainty because of internal variability) in many regions of the world (not shown). While this agreement between datasets cannot guarantee low observational uncertainty, as the datasets often share some in situ measurements, this adds confidence to constrained projections.

c. Where and when can we constrain?

We extend our analysis to other “Giorgi” regions (Giorgi and Francisco 2000) as well as larger custom regions (defined in Table S2). From a model perspective, the reduction in disagreement across the model response for the region at hand is feasible if the above-described criteria 1 and 2 are satisfied. We find that for today's warming level (roughly 0.25°C with respect to 1979–2010 from reanalyses), this is the case in CNA, the Amazon basin (AMZ), Greenland (GRL), north Asia

(NAS), and Australia (AUS)—regions highlighted in red and orange in Fig. 3. If we consider the historical TXx scaling from 1951 to 2014 (i.e., historical scaling based on years rather than level of warming), then Central America (CAM), central Asia (CAS), Alaska (ALA), and East Africa (EAF) can also be constrained because some models show more than the observed 0.25°C warming; therefore, more years are used to estimate the TXx scaling. If any parts of either criterion 1 or criterion 2 are not met, we estimate the additional future warming required to meet the criterion. The longer we wait, the stronger the signal and the more robust the TXx scaling will become. This provides an estimate of when we will be able to have a reasonably accurate historical TXx scaling. The timing of this depends on the region and is earlier for larger regions than smaller regions.

Even if criteria 1 and 2 are met, there are often no observations available (Zhang et al. 2011). Coverage in the observational datasets is limited, and the extension and refinement of observational datasets are hindered by several factors. A large part of the existing data of the twentieth century is being lost prior to getting digitized and included in observational datasets (Page et al. 2004; Donat et al. 2013a). At the same time, the lack of funding to maintain robust and dense observational networks leads to errors in the data as well as incomplete coverage (Munang et al. 2013; Alexander et al. 2006; Trenberth et al. 2013). Furthermore, some institutions are reluctant to share the already existing data with a wider community, confining the datasets even further. Therefore, we ask whether a constraint would be possible if observations were available. To this end, we use a perfect model approach to estimate by how much model disagreement could be reduced if we had perfect observational coverage since 1951 or even 1901. To achieve this, we treat individual models as observations (averaging multiple members if they exist) and quantify the resulting reduction in spread as a result of the constraint. We repeat this for all models available to get a reliable estimate. The improved agreement across models is quantified as a percent reduction in the range of projections (estimated as a standard deviation to be consistent) from the constrained ensemble with respect to the full ensemble (Fig. 4a). Two alternative methods were used to evaluate models relative to observations, but the width and magnitude of the constrained ensemble were found to be similar (Fig. S2). These alternative methods are described in the supplementary material.

We find that if observations were available since 1951, for 8 (5 Giorgi regions from Fig. 4a and 3 custom regions, not shown) out of 29 regions (21 Giorgi and 8 custom regions) the model disagreement could be reduced by about half in a perfect model approach. That is assuming

FIG. 3. Global mean warming level needed to constrain in a particular region. Hatching indicates regions where observations are available. Regions where the current method fails even if the full length of simulations (1951–2100) is considered (gray). Regions analyzed: AUS (45°–11°S, 110°–155°E), AMZ (20°S–12°N, 82°–34°W), SSA (56°–20°S, 76°–40°W), CAM (10°–30°N, 116°–83°W), western North America (WNA; 30°–60°N, 130°–103°W), CNA (30°–50°N, 103°–85°W), eastern North America (ENA; 25°–50°N, 85°–60°W), ALA (60°–72°N, 170°–103°W), GRL (50°–85°N, 103°–10°W), Mediterranean basin (MED; 30°–48°N, 10°W–40°E), northern Europe (NEU; 48°–75°N, 10°–40°W), western Africa (WAF; 12°S–18°N, 20°W–22°E), eastern Africa (EAF; 12°S–18°N, 22°–52°E), southern Africa (SAF; 35°–12°S, 10°W–52°E), Sahara (SAH; 18°–30°N, 20°W–65°E), SEA (11°S–20°N, 95°–155°E), East Asia (EAS; 20°–50°N, 100°–145°E), South Asia (SAS; 5°–30°N, 65°–100°E), CAS (30°–50°N, 40°–75°E), Tibet (TIB; 30°–50°N, 75°–100°E), and NAS (50°–70°N, 40°–180°E).

that the potential observations would not be far outside the model range. If observations were even available since 1901, then constraints would be possible over more regions (Fig. 4a). This emphasizes the importance of an effort to make all possible observational data available for as long a period as possible.

d. Case studies

Given the sparse observational coverage, only four Giorgi (Fig. 4a) and two custom regions (not shown) can in fact be evaluated against observations. For these regions observational constraints reduce the uncertainty in long-term TXx scaling. There is a slight model tendency to overestimate the observed historical TXx scaling over some of the regions (Fig. 4). Hence, the ensemble constrained on observations, excluding very high estimates of long-term TXx scaling, would imply very high future changes in the magnitude of TXx values. For example, we find two models—MIROC-ESM-CHEM and MIROC-ESM—with unrealistically high TXx values in dry regions exceeding 70°C in many parts of the world even in the historical period and the control simulations.

Constrained TXx scaling estimates often translate to narrower long-term projections of TXx at 3°C of global

mean warming (red box plots in Figs. 4b–e). We estimate long-term projections of TXx by multiplying TXx scaling with the corresponding local mean warming. Observational constraints suggest that TXx values are likely to warm slower than suggested by the multimodel estimate in Australia, north Asia, and Alaska (Figs. 4b,d,e), and faster than suggested in central North America. It is important to note that in the regions where local mean warming is uncertain, improved model agreement in TXx scalings does not translate to more certain long-term projections in TXx. An example of this is in Eurasia and Canada (Fig. S2).

4. Constraints on changes in frequency of hot extremes

a. Sources of model disagreement for future frequencies

In this section we show that the uncertainty in projected frequencies of hot extremes can be reduced in many parts of the world. The disagreement in future frequencies across models can be broken down into the following three contributions: (i) the inaccurate statistical representation of the present-day daily temperature distribution, (ii) model uncertainty in the local mean

FIG. 4. (a) Improvement in model agreement in TXx scaling estimates as a result of observational constraints in a perfect model framework. Results are expressed as the percent reduction in model spread relative to the full model spread. The top and bottom of the box in the boxplot visualize the first and third quartiles, respectively. The whiskers represent 1.5 times the interquartile range distance from the median. Giorgi regions were analyzed for periods starting in 1951 (pink) and 1901 (green) to constrain long-term TXx scaling estimates at 3°C global mean warming. Note that missing box plots mean that either criterion 1 or criterion 2 is not met in the present day or the target period. The lower panels show constrained projections for (b) Australia, (c) central North America, (d) north Asia, and (e) Alaska. Left axis (blue) corresponds to long-term TXx scaling at 3°C global mean warming. Right axis (red) shows the respective changes in TXx.

warming, consistent with a certain level of global mean, and (iii) changes in variance and higher-order moments of the temperature distribution with climate change. We aim to disentangle these contributions to identify regions where the shape of the present-day distribution [(i)],

which can be evaluated against observations, is the main source of model disagreement in projected frequencies. We do so by shifting the present-day temperature distribution by the local warming (multimodel mean or warming of the respective model). A comparison between

FIG. 5. A schematic diagram illustrating changes in frequency from a uniform shift of (a) wide and (b) narrow distributions.

estimated and simulated frequencies reveals the importance of these individual contributions [(i), (ii), or (iii)].

Assuming a uniform shift of the temperature distribution, models with a narrow temperature distribution (low variability; Fig. 5b) show a larger increase in frequencies for a given warming than models with wide temperature distributions (high variability; Fig. 5a) (Sillmann et al. 2014). To quantify the uncertainty contribution due to the representation of the present-day distribution [(i)], we shift the local daily temperature distribution of each model by the corresponding local multimodel mean warming consistent with 3°C global warming (see section 2b for details). The model disagreement in frequencies estimated from a multimodel mean shift is then entirely due to the differences in the model representation of the present-day distribution, as the shift is identical across models. In the Northern Hemisphere, we find a strong relationship between model-simulated frequencies and frequencies estimated from the multimodel mean shift (Fig. 6a). This implies that an accurate representation of the present-day temperature distribution is important. For the Northern Hemisphere, more than half of the model disagreement

(an explained variance of 0.6) in projected frequencies can be explained by the model's representation of the present-day temperature distribution.

However, for a given level of global mean warming of, for example, 3°C , models simulate different magnitudes of local warming for any given grid point. Assuming a simple shift of the present-day temperature distributions, this disagreement in local warming contributes to the uncertainties in simulated frequencies. To estimate this contribution [(ii)], we repeat the abovementioned analysis but shift the distribution at each grid point with each model's individual local warming rather than the multimodel mean warming. The difference between uncertainties in frequencies of hot extremes estimated from shifting with multimodel mean local warming and each model's individually simulated local mean warming is used as a quantification of the uncertainty contribution caused by model uncertainty in the local mean warming [(ii)]. This contribution explains an additional 30% of the disagreement in frequencies of hot extremes over the Northern Hemisphere (Fig. 6b). The remaining disagreement across models, which cannot be explained by (i) or (ii), can be attributed to the uncertain changes of

FIG. 6. Relationship between simulated and estimated frequencies from (a) multimodel mean warming and (b) single-model warming. Temperature shifts and warming magnitude present in the simulated frequencies are consistent with 3°C global mean warming. (c) Relationship between the 95-to-75th-percentile difference and the simulated frequencies. (d) As in (c), but adjusted for observational coverage. Observational estimates from HadGHCND, ERA-Interim, and NCEP-2 datasets (vertical lines). Models that have more than one ensemble member are color coded.

higher-order statistical moments of the daily temperature distribution, such as variance, skewness, and kurtosis of the temperature distribution [(iii)], that can be induced either by a systematic change (forced) or by internal variability (unforced). The difference between uncertainties in frequencies of hot extremes as directly simulated by the models and the ones estimated by uniformly shifting the temperature distribution with a model's local mean warming is used to quantify the contribution of changes in higher-order statistical moments.

In the regions where the uncertainty contribution induced by the inaccurate representation of the

present-day temperature distribution is large, the projected frequency of hot extremes can be constrained. These regions show a high correlation between estimated (from multimodel shift) and simulated frequencies (explained variance shown in Fig. 7a), even at a gridpoint scale. While a model evaluation is not performed at single grid points because of high internal variability, the maps (Fig. 7) provide a spatial perspective of the uncertainty contribution discussed above. The uncertainty caused by the model representation of the present-day distribution is dominant in many regions (Fig. 7a) except for North America, eastern Europe, and

a) Uncertainty due to the TX distribution

b) Uncertainty due to local warming (approx.)

c) Uncertainty due to changes in TX distribution (approx.)

FIG. 7. (a) Uncertainty caused by the representation of the present-day distribution in simulated frequencies at 3°C global mean warming expressed as explained variance. Approximate uncertainty due to (b) local warming and (c) changes of the temperature distribution expressed as explained variance. Only grid points where the contribution is significant ($p < 0.05$) are displayed. Models with multiple runs are averaged prior to computing the explained variance.

FIG. 8. Breakdown of individual uncertainty contributions in simulated frequencies for individual regions expressed as explained variance. Regions analyzed are AUS (45°S – 11°S , 110°E – 155°E), NH (0° – 90°N , 180°W – 180°E), central and eastern North America (CNA_ENA; 25° – 50°N , 103° – 60°W), and SEA (11°S – 20°N , 95° – 155°E). All correlations are calculated including only one run per model.

east Australia, where uncertainties caused by the local warming (Fig. 7b) are important. Also, changes in higher-order statistical moments are important at high latitudes (Fig. 7c), which makes it hard to constrain extremes using our approach. Based on this we break down the model disagreement in simulated frequencies into the individual contributions for selected regions (coordinates in Table S2). Figure 8 confirms that the representation of the present-day distribution is often the dominant source of uncertainty in projected frequencies of hot extremes at regional scales. Model disagreement in simulated frequencies can be reduced by as much as 85% in Southeast Asia (SEA) and 60% in Australia. It should be noted that the individual contributions are not independent and consequently the variances explained by individual factors often sum up to more than 100%.

b. Can we constrain changes in frequencies?

In regions where the representation of the present-day distribution accounts for more than half of the uncertainty (hatching in Fig. S3), we can substantially constrain projections. The temperature difference between the 95th and 75th percentiles of the present-day distribution (see section 2b) is our metric for evaluating the representation of the temperature distribution. In many regions, days that in the present day fall within the 95th–75th percentile temperature range are likely to exceed the present-day 95th percentile threshold at 3°C of global warming. Where observations and reanalyses exist, large discrepancies between datasets relative to model disagreement (Fig. S3c) hamper the strict evaluation of models. This points to large observational uncertainty in the daily data. The HadGHCND dataset may involve

inhomogeneities and uncertainties because of incomplete station density and gridding procedures. Nevertheless, it is still widely used to assess model performance (e.g., Donat and Alexander 2012; Sillmann et al. 2014; Perkins et al. 2012). We therefore do not strictly evaluate and constrain models but provide more general statements about the more likely range of projections. We find that relative to the HadGHCND, both reanalyses (Figs. S3e,f) and models (Fig. S3d) tend to overestimate the 95th-to-75th-percentile difference in most land regions, apart from North America. Since for a given warming wide distributions (high variability) tend to show lower future frequencies, this implies that many models are likely to underestimate the frequency increase in hot extremes (at least in the stippled areas in Fig. S3). This tendency to underestimate frequency increases is found on average over Northern Hemispheric land regions (Fig. 6c, compared to reanalyses over land-covered areas; Fig. 6d, compared to available datasets over land regions after accounting for observational availability). A similar conclusion holds for Australia (Fig. S4) and Southeast Asia (not shown).

5. Summary and conclusions

This study provides a systematic assessment of the potential to reduce model uncertainties in local projections of hot temperature extremes. We explore changes in both the intensity and frequency of hot extremes. We use observational constraints to link performance during the present-day and past climate (historical period) to long-term projections, and evaluate models against observations. To constrain projections

of the intensity of hot temperature extremes, we evaluate the historical relationship of annual temperature maxima (TXx) and mean summer temperatures (TXx scaling). Models with high historical TXx scaling tend to project a stronger future intensification of temperature extremes. To constrain projections of the frequencies of hot extremes (defined as future exceedance of today's 95th percentile), we evaluated the tail width of the present-day distribution of daily maximum temperatures. We show that in many parts of the world, a model's representation of the present-day temperature distribution is strongly related to its projection of frequency of hot extremes for a given level of warming. These relationships between historical characteristics of the temperature distribution and future changes are intuitive and easy to understand. To first order, an increase in both intensity and frequency of hot extremes is a consequence of shifting the temperature distribution. Model differences in the representation of the present-day distribution and the evolution of the temperature distribution in time determine the spread in projections. Present-day biases in the representation of the distribution and uncertain changes in the temperature distribution have different importance for projections of intensities as opposed to frequencies. For example, a present-day bias in the width of the distribution is more important for projections of frequencies of hot extremes than for projections of intensities. This is why the two aspects are studied separately.

Based on observational constraints, we show that on average climate models likely underestimate the increase in frequencies over a large portion of the Northern Hemisphere and Australia for a given warming. The main reason is that models tend to simulate a too-wide-tailed (too variable) present-day temperature distribution. We also find that uncertainties in projections of intensities of hot extremes can be significantly reduced in several regions, including Australia, central North America, and north Asia. In contrast to underestimated frequencies, models were found to likely overestimate future intensities of hot temperature extremes for a certain level of warming over these regions, a tendency of models that is also evident in simulations of the present day (e.g., Yao et al. 2013; Morak et al. 2013; Kjellström et al. 2007).

We identify two main factors that make observational constraints for hot extremes challenging: (i) the effect of internal unforced variability and (ii) the lack of long, reliable observational records with dense global coverage. The latter are more limiting for projections of frequencies of hot extremes, as not only indices of annual maxima but actual daily data are required. The availability of extreme indices (e.g., TXx) is more complete

than for raw daily temperature data, but it still suffers from poor observational coverage and quality, inhibiting the evaluation of models in many regions where the method would be applicable otherwise. We demonstrated in a perfect model approach that if observations in these regions were available since 1951 or even 1901, then the uncertainty in projections of the intensification of hot extremes could be reduced by about half. This emphasizes the importance of long-term and high-quality observational networks. For instance, in the tropical regions, the projections of both the amplitude and frequency of hot temperature extremes could be easily constrained if observations existed. This would be beneficial for local communities, as they tend to be highly vulnerable to climate change, in particular to changes in extremes (UNFCCC 2007). The fact that in many cases observational constraints could narrow uncertainties but are hindered by observational data availability stresses the urge for immediate action to improve the availability of and access to observational networks.

The role of internal variability is particularly important for projections of changes in intensity of temperature extremes. There are several studies that evaluate models on how well they reproduce extremes (Kharin et al. 2013; Perkins et al. 2007; Sillmann et al. 2013a), but the role of internal variability is often not accounted for, as only one realization per model is analyzed. In this study, we showed that internal variability can prevent robust model evaluation, particularly if the period analyzed is short. Internal variability is less problematic for projections of frequencies of hot extremes. This is because the large uncertainty in the shape of the very tail that dominates in projections of changes in intensity becomes less relevant for the frequency once much of the distribution has exceeded the threshold.

In principle, the approach of observational constraints can be applied to other variables (e.g., precipitation) and a wider range of percentiles (e.g., 90th percentile of annual daily temperatures or the coldest annual temperatures TNn). Nevertheless, all criteria and limitations listed in this paper also apply to these other variables. If available, then observations should be used to pin down the range of plausible projections, as a wide range of model responses contributes to a cascade of uncertainties in impact models. Reducing uncertainties in projections is beneficial for the scientific communities that study climate change impacts. Also, better knowledge about what features of the present-day climate are important for skillful projections is imperative for the model development community.

Acknowledgments. We acknowledge the World Climate Research Programme's Working Group on Coupled

Modelling, which is responsible for CMIP. We thank the climate modeling groups (listed in Table S1) for producing and making available their model output. Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung provided funding for this study. 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.

REFERENCES

- Alexander, L. V., 2016: Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Wea. Climate Extremes*, **11**, 4–16, doi:[10.1016/j.wace.2015.10.007](https://doi.org/10.1016/j.wace.2015.10.007).
- , and J. M. Arblaster, 2009: Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *Int. J. Climatol.*, **29**, 417–435, doi:[10.1002/joc.1730](https://doi.org/10.1002/joc.1730).
- , and C. Tebaldi, 2012: Climate and weather extremes: Observations, modeling, and projections. *The Future of the World's Climate*, 2nd ed. A. Henderson-Sellers and K. McGuffie, Eds., Elsevier, 253–288, doi:[10.1016/b978-0-12-386917-3.00010-5](https://doi.org/10.1016/b978-0-12-386917-3.00010-5).
- , and S. Perkins, 2013: Debate heating up over changes in climate variability. *Environ. Res. Lett.*, **8**, 041001, doi:[10.1088/1748-9326/8/4/041001](https://doi.org/10.1088/1748-9326/8/4/041001).
- , and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi:[10.1029/2005JD006290](https://doi.org/10.1029/2005JD006290).
- Ballester, J., F. Giorgi, and X. Rodó, 2010: Changes in European temperature extremes can be predicted from changes in PDF central statistics. *Climatic Change*, **98**, 277–284, doi:[10.1007/s10584-009-9758-0](https://doi.org/10.1007/s10584-009-9758-0).
- Boberg, F., and J. H. Christensen, 2012: Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nat. Climate Change*, **2**, 433–436, doi:[10.1038/nclimate1454](https://doi.org/10.1038/nclimate1454).
- Boé, J., A. Hall, and X. Qu, 2009: September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nat. Geosci.*, **2**, 341–343, doi:[10.1038/ngeo467](https://doi.org/10.1038/ngeo467).
- Borodina, A., E. M. Fischer, and R. Knutti, 2017: Emergent constraints in climate projections: A case study of changes in high latitude temperature variability. *J. Climate*, **30**, 3655–3670, doi:[10.1175/JCLI-D-16-0662.1](https://doi.org/10.1175/JCLI-D-16-0662.1).
- Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, **111**, D05101, doi:[10.1029/2005JD006280](https://doi.org/10.1029/2005JD006280).
- Caldwell, P. M., C. S. Bretherton, M. D. Zelinka, S. A. Klein, B. D. Santer, and B. M. Sanderson, 2014: Statistical significance of climate sensitivity predictors obtained by data mining. *Geophys. Res. Lett.*, **41**, 1803–1808, doi:[10.1002/2014GL059205](https://doi.org/10.1002/2014GL059205).
- Cattiaux, J., H. Douville, and Y. Peings, 2013: European temperatures in CMIP5: Origins of present-day biases and future uncertainties. *Climate Dyn.*, **41**, 2889–2907, doi:[10.1007/s00382-013-1731-y](https://doi.org/10.1007/s00382-013-1731-y).
- Christensen, J. H., and F. Boberg, 2012: Temperature dependent climate projection deficiencies in CMIP5 models. *Geophys. Res. Lett.*, **39**, L24705, doi:[10.1029/2012GL053650](https://doi.org/10.1029/2012GL053650).
- Christidis, N., P. A. Stott, S. Brown, G. C. Hegerl, and J. Caesar, 2005: Detection of changes in temperature extremes during the second half of the 20th century. *Geophys. Res. Lett.*, **32**, L20716, doi:[10.1029/2005GL023885](https://doi.org/10.1029/2005GL023885).
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:[10.1002/qj.776](https://doi.org/10.1002/qj.776).
- Coumou, D., and A. Robinson, 2013: Historic and future increase in the global land area affected by monthly heat extremes. *Environ. Res. Lett.*, **8**, 034018, doi:[10.1088/1748-9326/8/3/034018](https://doi.org/10.1088/1748-9326/8/3/034018).
- , —, and S. Rahmstorf, 2013: Global increase in record-breaking monthly-mean temperatures. *Climatic Change*, **118**, 771–782, doi:[10.1007/s10584-012-0668-1](https://doi.org/10.1007/s10584-012-0668-1).
- Cowan, T., A. Purich, S. Perkins, A. Pezza, G. Boschat, and K. Sadler, 2014: More frequent, longer, and hotter heat waves for Australia in the twenty-first century. *J. Climate*, **27**, 5851–5871, doi:[10.1175/JCLI-D-14-00092.1](https://doi.org/10.1175/JCLI-D-14-00092.1).
- Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke, 2013: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, **494**, 341–344, doi:[10.1038/nature11882](https://doi.org/10.1038/nature11882).
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- Della-Marta, P. M., M. R. Haylock, J. Luterbacher, and H. Wanner, 2007: Doubled length of western European summer heat waves since 1880. *J. Geophys. Res.*, **112**, D15103, doi:[10.1029/2007JD008510](https://doi.org/10.1029/2007JD008510).
- Donat, M. G., and L. V. Alexander, 2012: The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.*, **39**, L14707, doi:[10.1029/2012GL052459](https://doi.org/10.1029/2012GL052459).
- , and Coauthors, 2013a: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, doi:[10.1002/jgrd.50150](https://doi.org/10.1002/jgrd.50150).
- , L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013b: Global land-based datasets for monitoring climatic extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, doi:[10.1175/BAMS-D-12-00109.1](https://doi.org/10.1175/BAMS-D-12-00109.1).
- Fasullo, J. T., and K. E. Trenberth, 2012: A less cloudy future: The role of subtropical subsidence in climate sensitivity. *Science*, **338**, 792–794, doi:[10.1126/science.1227465](https://doi.org/10.1126/science.1227465).
- Fischer, E. M., and C. Schär, 2009: Future changes in daily summer temperature variability: Driving processes and role for temperature extremes. *Climate Dyn.*, **33**, 917–935, doi:[10.1007/s00382-008-0473-8](https://doi.org/10.1007/s00382-008-0473-8).
- , and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, doi:[10.1002/2013GL058499](https://doi.org/10.1002/2013GL058499).
- , and —, 2015: Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Climate Change*, **5**, 560–564, doi:[10.1038/nclimate2617](https://doi.org/10.1038/nclimate2617).
- , D. M. Lawrence, and B. M. Sanderson, 2011: Quantifying uncertainties in projections of extremes—A perturbed land surface parameter experiment. *Climate Dyn.*, **37**, 1381–1398, doi:[10.1007/s00382-010-0915-y](https://doi.org/10.1007/s00382-010-0915-y).
- , J. Rajczak, and C. Schär, 2012: Changes in European summer temperature variability revisited. *Geophys. Res. Lett.*, **39**, L19702, doi:[10.1029/2012GL052730](https://doi.org/10.1029/2012GL052730).

- , U. Beyerle, and R. Knutti, 2013: Robust spatially aggregated projections of climate extremes. *Nat. Climate Change*, **3**, 1033–1038, doi:[10.1038/nclimate2051](https://doi.org/10.1038/nclimate2051).
- Giorgi, F., and R. Francisco, 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182, doi:[10.1007/PL00013733](https://doi.org/10.1007/PL00013733).
- Gregory, J. M., and J. F. B. Mitchell, 1995: Simulation of daily variability of surface temperature and precipitation over Europe in the current and $2 \times \text{CO}_2$ climates using the UKMO climate model. *Quart. J. Roy. Meteor. Soc.*, **121**, 1451–1476, doi:[10.1002/qj.49712152611](https://doi.org/10.1002/qj.49712152611).
- Griffiths, G. M., and Coauthors, 2005: Change in mean temperature as a predictor of extreme temperature change in the Asia-Pacific region. *Int. J. Climatol.*, **25**, 1301–1330, doi:[10.1002/joc.1194](https://doi.org/10.1002/joc.1194).
- Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, L03502, doi:[10.1029/2005GL025127](https://doi.org/10.1029/2005GL025127).
- Holland, M. M., and C. M. Bitz, 2003: Polar amplification of climate change in coupled models. *Climate Dyn.*, **21**, 221–232, doi:[10.1007/s00382-003-0332-6](https://doi.org/10.1007/s00382-003-0332-6), doi:[10.1029/2005GL025127](https://doi.org/10.1029/2005GL025127).
- Holmes, C. R., T. Woollings, E. Hawkins, and H. de Vries, 2016: Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Climate*, **29**, 2221–2236, doi:[10.1175/JCLI-D-14-00735.1](https://doi.org/10.1175/JCLI-D-14-00735.1).
- Huber, M., I. Mahlstein, M. Wild, J. Fasullo, and R. Knutti, 2011: Constraints on climate sensitivity from radiation patterns in climate models. *J. Climate*, **24**, 1034–1052, doi:[10.1175/2010JCLI3403.1](https://doi.org/10.1175/2010JCLI3403.1).
- Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223, doi:[10.1175/1520-0442\(2003\)016<0206:HALSSA.2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0206:HALSSA.2.0.CO;2).
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, doi:[10.1175/1520-0493\(1999\)127<2204:FASOCR.2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR.2.0.CO;2).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, doi:[10.1175/BAMS-83-11-1631](https://doi.org/10.1175/BAMS-83-11-1631).
- Katz, W. R., and B. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21**, 289–302, doi:[10.1007/BF00139728](https://doi.org/10.1007/BF00139728).
- Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate*, **13**, 3760–3788, doi:[10.1175/1520-0442\(2000\)013<3760:CITEIA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2).
- , and —, 2005: Estimating extremes in transient climate change simulations. *J. Climate*, **18**, 1156–1173, doi:[10.1175/JCLI3320.1](https://doi.org/10.1175/JCLI3320.1).
- , —, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, doi:[10.1175/JCLI4066.1](https://doi.org/10.1175/JCLI4066.1).
- , —, —, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, doi:[10.1007/s10584-013-0705-8](https://doi.org/10.1007/s10584-013-0705-8).
- Kjellström, E., L. Bärring, D. Jacob, R. Jones, G. Lenderink, and C. Schär, 2007: Modelling daily temperature extremes: Recent climate and future changes over Europe. *Climatic Change*, **81**, 249–265, doi:[10.1007/s10584-006-9220-5](https://doi.org/10.1007/s10584-006-9220-5).
- Klein Tank, A. M. G., and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *J. Climate*, **16**, 3665–3680, doi:[10.1175/1520-0442\(2003\)016<3665:TIHODT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3665:TIHODT>2.0.CO;2).
- Knutti, R., and L. Tomassini, 2008: Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett.*, **35**, L09701, doi:[10.1029/2007GL032904](https://doi.org/10.1029/2007GL032904).
- , G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Climate*, **19**, 4224–4233, doi:[10.1175/JCLI3865.1](https://doi.org/10.1175/JCLI3865.1).
- , J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, doi:[10.1002/2016GL072012](https://doi.org/10.1002/2016GL072012).
- Loeb, N. G., H. Wang, A. Cheng, S. Kato, J. T. Fasullo, K.-M. Xu, and R. P. Allan, 2016: Observational constraints on atmospheric and oceanic cross-equatorial heat transports: Revisiting the precipitation asymmetry problem in climate models. *Climate Dyn.*, **46**, 3239–3257, doi:[10.1007/s00382-015-2766-z](https://doi.org/10.1007/s00382-015-2766-z).
- Lovejoy, S., 2014: Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming. *Climate Dyn.*, **42**, 2339–2351, doi:[10.1007/s00382-014-2128-2](https://doi.org/10.1007/s00382-014-2128-2).
- Mahlstein, I., and R. Knutti, 2012: September Arctic sea ice predicted to disappear near 2°C global warming above present. *J. Geophys. Res.*, **117**, D06104, doi:[10.1029/2011JD016709](https://doi.org/10.1029/2011JD016709).
- Massonnet, F., T. Fichet, H. Goosse, C. M. Bitz, G. Philippon-Berthier, M. M. Holland, and P.-Y. Barriat, 2012: Constraining projections of summer Arctic sea ice. *Cryosphere*, **6**, 1383–1394, doi:[10.5194/tc-6-1383-2012](https://doi.org/10.5194/tc-6-1383-2012).
- Meehl, G. A., and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305**, 994–997, doi:[10.1126/science.1098704](https://doi.org/10.1126/science.1098704).
- Morak, S., G. C. Hegerl, and N. Christidis, 2013: Detectable changes in the frequency of temperature extremes. *J. Climate*, **26**, 1561–1574, doi:[10.1175/JCLI-D-11-00678.1](https://doi.org/10.1175/JCLI-D-11-00678.1).
- Munang, R., J. N. Nkem, and Z. Han, 2013: Using data digitalization to inform climate change adaptation policy: Informing the future using the present. *Wea. Climate Extremes*, **1**, 17–18, doi:[10.1016/j.wace.2013.07.001](https://doi.org/10.1016/j.wace.2013.07.001).
- Murray, V., and K. L. Ebi, 2012: IPCC special report on managing the risks of extreme events and disasters to advance climate change adaptation (SREX). *J. Epidemiol. Community Health*, **66**, 759–760, doi:[10.1136/jech-2012-201045](https://doi.org/10.1136/jech-2012-201045).
- O’Gorman, P., 2012: Sensitivity of tropical precipitation extremes to climate change. *Nat. Geosci.*, **5**, 697–700, doi:[10.1038/ngeo1568](https://doi.org/10.1038/ngeo1568).
- Page, C. M., and Coauthors, 2004: Data rescue in the Southeast Asia and South Pacific region. *Bull. Amer. Meteor. Soc.*, **85**, 1483–1489, doi:[10.1175/BAMS-85-10-1483](https://doi.org/10.1175/BAMS-85-10-1483).
- Perkins, S. E., and E. M. Fischer, 2013: The usefulness of different realizations for the model evaluation of regional trends in heat waves. *Geophys. Res. Lett.*, **40**, 5793–5797, doi:[10.1002/2013GL057833](https://doi.org/10.1002/2013GL057833).
- , A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007: Evaluation of the AR4 climate models’ simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate*, **20**, 4356–4376, doi:[10.1175/JCLI4253.1](https://doi.org/10.1175/JCLI4253.1).

- , L. V. Alexander, and J. R. Nairn, 2012: Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophys. Res. Lett.*, **39**, L20714, doi:[10.1029/2012GL053361](https://doi.org/10.1029/2012GL053361).
- Perkins-Kirkpatrick, S. E., E. M. Fischer, O. Angélil, and P. B. Gibson, 2017: The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.*, **12**, 44005, doi:[10.1088/1748-9326/aa63fe](https://doi.org/10.1088/1748-9326/aa63fe).
- Robeson, S. M., 2004: Trends in time-varying percentiles of daily minimum and maximum temperature over North America. *Geophys. Res. Lett.*, **31**, L04203, doi:[10.1029/2003GL019019](https://doi.org/10.1029/2003GL019019).
- Schaller, N., I. Mahlstein, J. Cermak, and R. Knutti, 2011: Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *J. Geophys. Res.*, **116**, D10118, doi:[10.1029/2010JD014963](https://doi.org/10.1029/2010JD014963).
- Schär, C., L. Vidale Pier, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, doi:[10.1038/nature02300](https://doi.org/10.1038/nature02300).
- Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land-atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–209, doi:[10.1038/nature05095](https://doi.org/10.1038/nature05095).
- , M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby, 2016: Allowable CO₂ emissions based on regional and impact-related climate targets. *Nature*, **529**, 477–483, doi:[10.1038/nature16542](https://doi.org/10.1038/nature16542).
- Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, doi:[10.1002/jgrd.50203](https://doi.org/10.1002/jgrd.50203).
- , —, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.*, **118**, 2473–2493, doi:[10.1002/jgrd.50188](https://doi.org/10.1002/jgrd.50188).
- , —, —, —, —, and M. G. Donat, 2014: Evaluating model-simulated variability in temperature extremes using modified percentile indices. *Int. J. Climatol.*, **34**, 3304–3311, doi:[10.1002/joc.3899](https://doi.org/10.1002/joc.3899).
- Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2011: Evolution of extreme temperatures in a warming climate. *Geophys. Res. Lett.*, **38**, L16701, doi:[10.1029/2011GL048437](https://doi.org/10.1029/2011GL048437).
- Son, S. W., and Coauthors, 2010: Impact of stratospheric ozone on Southern Hemisphere circulation change: A multimodel assessment. *J. Geophys. Res.*, **115**, D00M07, doi:[10.1029/2010JD014271](https://doi.org/10.1029/2010JD014271).
- Stegehuis, A. I., A. J. Teuling, P. Ciais, R. Vautard, and M. Jung, 2013: Future European temperature change uncertainties reduced by using land heat flux observations. *Geophys. Res. Lett.*, **40**, 2242–2245, doi:[10.1002/grl.50404](https://doi.org/10.1002/grl.50404).
- Su, H., J. H. Jiang, C. Zhai, T. J. Shen, J. D. Neelin, G. L. Stephens, and Y. L. Yung, 2014: Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *J. Geophys. Res. Atmos.*, **119**, 5787–5805, doi:[10.1002/2014JD021642](https://doi.org/10.1002/2014JD021642).
- Tian, B., 2015: Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophys. Res. Lett.*, **42**, 4133–4141, doi:[10.1002/2015GL064119](https://doi.org/10.1002/2015GL064119).
- Trenberth, K. E., and Coauthors, 2013: Challenges of a sustained climate observing system. *Climate Science for Serving Society*, G. R. Asrar and J. W. Hurrell, Eds., Springer, 13–50, doi:[10.1007/978-94-007-6692-1_2](https://doi.org/10.1007/978-94-007-6692-1_2).
- UNFCCC, 2007: Climate change: Impacts, vulnerabilities and adaptation in developing countries. UNFCCC Publ., 64 pp., <https://unfccc.int/resource/docs/publications/impacts.pdf>.
- Vogel, M. M., R. Orth, F. Cheruy, S. Hagemann, R. Lorenz, B. J. J. M. van den Hurk, and S. I. Seneviratne, 2017: Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. *Geophys. Res. Lett.*, **44**, 1511–1519, doi:[10.1002/2016GL071235](https://doi.org/10.1002/2016GL071235).
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp., doi:[10.1017/cbo9780511612336.018](https://doi.org/10.1017/cbo9780511612336.018).
- Wenzel, S., V. Eyring, E. P. Gerber, and A. Y. Karpechko, 2016: Constraining future summer austral jet stream positions in the CMIP5 ensemble by process-oriented multiple diagnostic regression. *J. Climate*, **29**, 673–687, doi:[10.1175/JCLI-D-15-0412.1](https://doi.org/10.1175/JCLI-D-15-0412.1).
- Yao, Y., Y. Luo, J. Huang, and Z. Zhao, 2013: Comparison of monthly temperature extremes simulated by CMIP3 and CMIP5 models. *J. Climate*, **26**, 7692–7707, doi:[10.1175/JCLI-D-12-00560.1](https://doi.org/10.1175/JCLI-D-12-00560.1).
- Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, doi:[10.1002/wcc.147](https://doi.org/10.1002/wcc.147).