Dear editor

We have considered all the major and minor comments by the reviewers and revised the manuscript accordingly. Following relevant changes were done to the manuscript:

- More accurate bias correction of the absolute values for daily maximum temperature (TX) using quantile-mapping (shown in updated Figure 4a)
- The bias corrected TX and hence, values for percentage of AgPop region affected by extremely hot temperatures, have changed due to the new bias correction method
- Added verification of bias corrected TX for 2018 using observational data sets and reanalysis (e.g. Fig. A3)
- Inclusion of the Eastern Asia SREX region in the figures and throughout the text
- Added discussion of model biases and uncertainties in observations of regional TX trends
- New Table A1 in the appendix showing observed regional TS trends and their uncertainties using two global mean temperature data sets (GISTEMPv4, HadCRUT4) and two regional land temperature data sets (CRU TSv4.03, Berkeley Earth)
- Included observed TX trends in Fig. 7
- Updated Fig. 7 also to show the study time period (13-27 July) for the nudged CESM simulations
- Clarification of the choice for the end date for the study period and the nudging input files as well as expanded discussion of the implications for regions that were affected by heat waves after this end date

Please find below a point-by-point reply to the comments made by the reviewers and a marked up manuscript version showing the changes made.

Sincerely
Kathrin Wehrli (on behalf of all authors)

Response to Anonymous Referee #1

The concept of asking how a given meteorological event might have been exacerbated by global warming has been growing in popularity as an alternative to the more common approach of probabilistic event attribution. As first proposed by Trenberth et al.(2015 doi: 10.1038/NCLIMATE2657) and Shepherd (2016 doi: 10.1007/s40641-016-0033-y), this 'storyline' approach takes the atmospheric flow configuration leading to the event as given, and quantifies the impact of global warming conditional on that flow configuration. The arguments for why this may be useful are given in those two papers,but the general concept is that generality is sacrificed in order to obtain a more detailed and hopefully more informative statement of impact (since tied to a particular event).The storyline approach was first applied to synoptic-timescale weather phenomena, e.g. tropical cyclones, where the conditionality was applied either through the initial conditions in a forecast, or through the boundary conditions in a regional model. Here the storyline approach is applied to a multi-week heat wave event, nudging the circulation in a global model to reanalysis, following the methodology previously used by the authors in their 2019 JGR paper to understand the role of soil-moisture feedbacks in heat waves. It is important to document applications of the storyline approach in different contexts so that we can learn to understand its strengths and weaknesses. From that perspective this study is welcome, and for the most part the results are carefully explained and clearly presented. I find Figure 7 to be the most interesting of all. I am happy to recommend publication, provided the following points are addressed:

A1: We thank the reviewer for the positive and thoughtful evaluation of the manuscript. We appreciate the comments on the selection of the study regions and the bias correction of absolute maximum daily temperatures. We have followed the recommendations and computed new figures that will be included and discussed in the revised manuscript. We now use quantile-mapping to bias-correct our model simulations. The area affected by maximum daily temperature > 40°C has changed due to the new bias correction method. The new results agree better with observations and are qualitatively still in-line with earlier results. Below we will answer the specific questions of the reviewer.

1. Figure 3 shows only temperature anomalies. It would be good to also show absolute temperatures (e.g. in the maps), so that the reader can see the extent of the temperature bias of the model.

A2: We agree with the reviewer that biases of absolute temperature of the model and their correction is important. It is well-known that the majority of CMIP5 models, including CESM, overestimate summer temperatures in Northern Hemisphere midlatitudes (e.g. Mueller and Seneviratne, 2014, GRL; Wehrli et al., 2018, GRL; also shown for TXx in CESM in the latter). In the revised manuscript we add plots of absolute TX (mean over Jul. 13-27 2018) to Appendix Figure A3, which show the bias-corrected (using quantile mapping) CESM historical simulation against different references. Since it is known that model biases are large, we do not think it is necessary to show the magnitude of the bias. It is, however, crucial to discuss the bias correction and check the bias-corrected model against reference data sets. For reference we show a comparison of TXx > 40°C for the original model output, mean bias correction and quantile mapping (using both Berkeley-Earth and ERA-Interim as reference) against ERA-Interim and Berkeley Earth (see attachment/below).

References:

Mueller, B., and Seneviratne, S. I. (2014), Systematic land climate and evapotranspiration biases in CMIP5 simulations, *Geophys. Res. Lett.*, 41, 128-134, doi:10.1002/2013GL058055.

Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., & Seneviratne, S. I. (2018). Assessing the dynamic versus thermodynamic origin of climate model biases. *Geophysical Research Letters*, 45, 8471-8479. https://doi.org/10.1029/2018GL079220

2. In some periods and regions, the differences between the nudged run and ERA-Interim anomalies in the time series in Figure 3 can exceed 1°C for extended periods. Do you have any idea why this would be the case, given that generally the differences are much smaller?

A3: We can only speculate about the reason for the large differences between the nudged run and ERA-Interim temperature anomalies in Figure 3. The most striking case is NEU and we verified that the difference is largest during two shorter periods in mid- and end of June and the period shown in Figure 3 in the beginning of July. For these periods warm anomalies are overestimated, whereas during the rest of the year the differences are generally smaller. One possibility is that during the NH2018 heatwave soil moisture got depleted even in regions that are usually rich in moisture (so-called wet regime), which causes the land surface to react very sensitive to a further decrease in soil moisture and to incoming radiation (i.e. change from wet to transitional regime). If the model dries faster or transitions to a radiation-sensitive state earlier than ERA-Interim this might result in a more sensitive and more pronounced response in temperature. With decreasing moisture availability more incoming radiation will contribute to sensible heat flux and hence to increased temperature.

3. Comparing Figure 1 and Figure 3, with the exception of the southern portion of NEU the study areas seem almost to be orthogonal to the areas of maximum temperature anomaly, and one of the most striking AgPop regions where there is a high temperature anomaly, eastern Asia, is not included in the study. Thus the choice of study areas seems quite odd. It would surely be straightforward to include a relevant east Asian SREX region for completeness, which would mitigate the European/North American bias of this study.

A4: We agree with the reviewer that our choice of regions was biased towards Europe and North America. In the revised manuscript we include the Eastern Asian SREX region (EAS) in the figures and analysis. The region of Neufundland/Québec would also be interesting to examine. However, there is no SREX region that would be suitable. The Canada/ Greenland/ Iceland SREX region (CGI) encompasses large areas with temperature anomalies of the opposite sign (e.g. Greenland). We decided to not define a new region specifically for this case.
Apart from the just-mentioned region in north-eastern America, we believe that the interesting regions for the 2018 heat wave are addressed in this study. The Mediterranean was not strongly affected by heat waves during summer 2018, which was discussed in other studies (e.g. Toreti et al., 2019). Therefore, we thought it is interesting to include this region in the analysis. Except for the scaling plots (Figure 7) and time series in Figure 3, results always show the entire Northern Hemisphere north of 25°N and the discussion is not limited to the SREX regions.

Reference:
Toreti, A., Belward, A.,Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie, O., et al. (2019). The exceptional 2018 European water seesaw calls for action on adaptation. *Earth's Future*, 7, 652–663. https://doi.org/10.1029/2019EF001170

4. In all three SREX regions of North America, the difference between the nudged run and either ERA-Interim or Berkeley for the maximum daily temperature anomalies (Figure A4), especially for some of the largest values, can be much larger than the difference of the mean daily temperature anomalies (Figure 3). What is the reason for that? And how does it affect

your estimates of extreme temperature? This feature suggests that using the climatological mean TX to bias-correct the TX values may not be adequate.

A5: One reason why differences of maximum temperature are larger than differences of mean daily temperature is that in the latter biases during the night can be balanced out by biases during the day and vice versa (overestimated/underestimated diurnal cycle can still show as correct daily mean). Connected to that, biases of mean daily temperature anomalies can affect biases of extreme temperatures (and vice versa) but they don't have to in a direct, linear way. We agree that the bias-correction of TX should be treated in more details and several methods should be evaluated for their adequacy. In the revised manuscript we apply a quantile-mapping that was tested for two reference data sets (see next answer A6).

5. The left column of Figure 4 apparently includes a bias correction of the model output.This is only mentioned in the figure caption, not in the methods or anywhere else. Since the bias correction is almost certain to affect the results of the study, which are framed relative to a fixed temperature threshold of 40°C, a much more detailed assessment of its effect, and the potential error incurred thereby, is required. It appears that the bias correction was simply an adjustment of the mean, which assumes that the model TX distribution is perfect. Can you support this assumption with evidence? As noted in the previous comment, the assumption would appear to be contradicted by your own results. Why did you not use quantile mapping or some other more detailed method,which would treat the tails differently from the mean?

A6: We agree with the reviewer that the bias correction was not addressed sufficiently in the manuscript. The presented method was a day-of-year dependent correction of the mean (TX) bias. We tested the presented method against a quantile-mapping bias correction using a 91-day moving window (hence, making it also dependent on the day-of-year). As could be expected, the quantile-mapped results reveal that the bias correction method strongly influences the results and the mean bias correction is less appropriate in our case. We discuss this in the revised manuscript. Figure 4 and the numbers in the manuscript are replaced by the results from the quantile-mapping. We verified that the quantile mapping leads to better results: TX RMSE for the study period (Jul. 13-27 2018) and all land areas north of 25°N is reduced from 7.48°C in the original model to 1.95°C using mean bias correction and to 1.45°C using quantile mapping (Berkeley-Earth as reference; qualitatively the same is true for the RMSE of the AgPop region). The area affected by temperatures > 40°C discussed in the conclusion better match the reference data sets (Berkeley, ERA-Interim, MERRA-2) when using the quantile mapping. Below we included a figure (Figure 1) to show the effect of bias correction on TXx. The values for the AgPop region > 40°C are 9.1% and 8.5% for ERA-Interim and Berkeley-Earth, respectively. The bias corrected model simulates 8.8% and 9.3% area affected (depending on the reference used for the calibration of the quantile mapping), which is more accurate than the 20% we obtain for the mean bias correction.

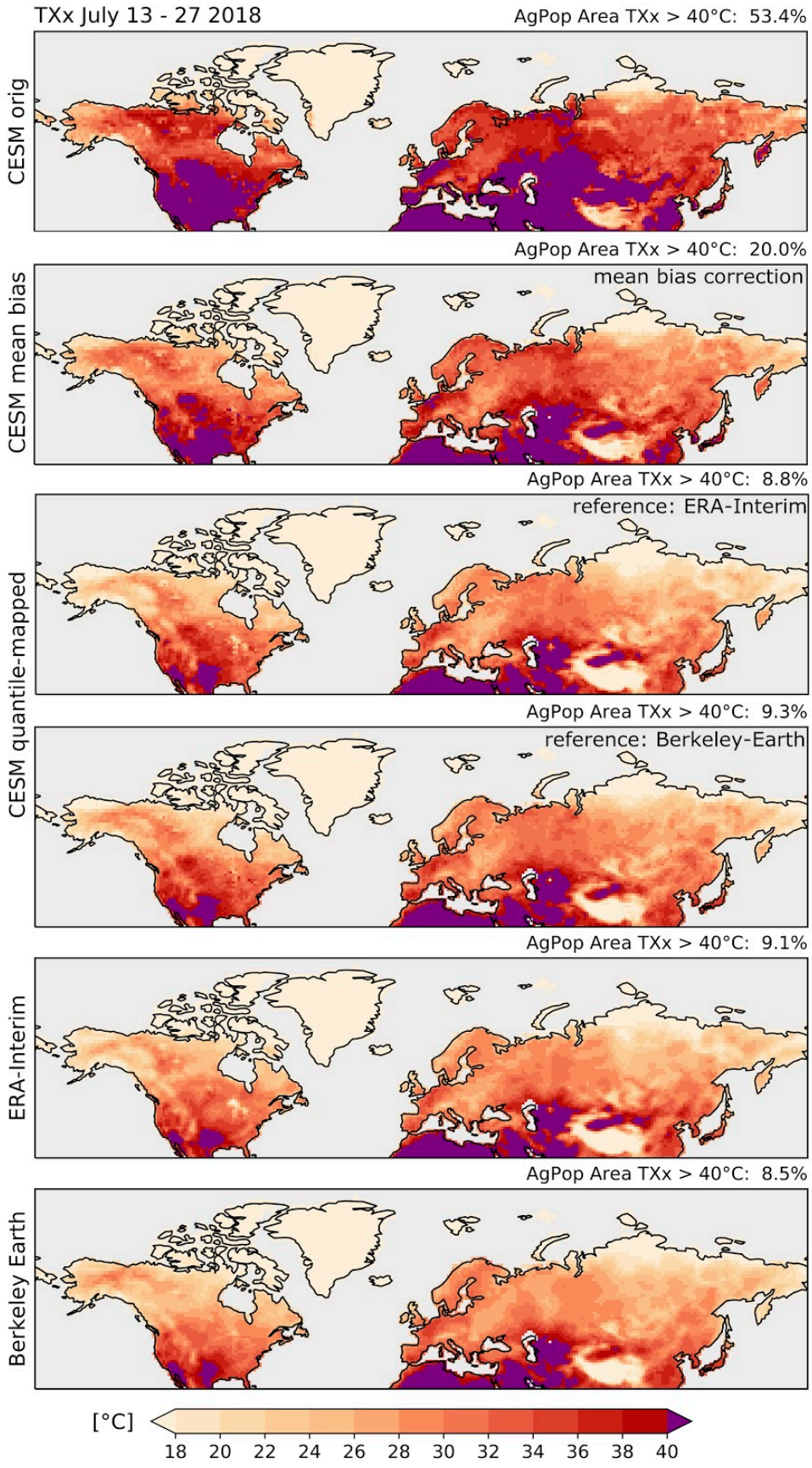

Figure 1: TXx for 13-27 July 2018 and fraction of the AgPop region experiencing maximum daily temperatures larger than 40°C for the original and bias-corrected model output as well as two reference data sets.

6. In lines 9-10, you should mention also the percentage value for the actual event, as a reference.

A7: We agree and adjusted the last line of the abstract to include the percentage value for the actual event (now changed to the new value from the quantile-mapped simulations).

Response to Geert Jan van Oldenborgh

This paper analyses what the heat waves of the NH summer of 2018 would look like in pre-industrial, current, 1.5°C, 2°C, 3°C and 4°C climates with the same circulation prescribed. This prescription makes it impossible to estimate probabilities, but the dependence of local heat wave temperatures and other properties such as drought and solar radiation can be shown, and a scaling with the global mean temperature established. The results look solid and are certainly interesting, giving the important message that local temperature effects may be much stronger than the global mean temperature rise. It does not address the possible interconnectedness of the heatwaves around the globe in that summer beyond citing Kornhuber (2019b). There is only one major comment I have on the analysis, namely that it is only analyses climate model data and does not make any connection to observations beyond showing the patterns agree. Trends in heat waves are notoriously badly simulated by climate models and some comparisons of the modelled trends to the observed trends would make the paper and a discussion on possible discrepancies and how these would affect future trends much more useful for readers who want to apply the results to the real world rather than the model world. As an example, I computed the observed trends corresponding to Fig.7 from CRU TS4.04. The observed scaling factors are very different from the modelled ones, lower in North America and higher in Europe:
WNA: 0.9±0.2 K/K, CNA: 0.2±0.3 K/K (see eg https://doi.org/10.1038/s41467-020-16676-w), ENA: 0.6±0.2 K/K.
NEU: 1.2±0.3 K/K, CEU: 1.5±0.3 K/K, MED: 1.8±0.2 K/K.
With the addition of observed trends and a discussion on the differences with climate models (and the minor comments below) the paper would be a useful contribution to the literature.

B1: We thank Geert Jan van Oldenborgh for his comment on observed trends and agree that it should be discussed in the paper. Therefore, we made additional analyses using maximum daily temperatures for CRU TS4.03 and the Berkeley Earth Surface Temperature (BEST) project. As a reference for global mean temperature change (land+ocean) we used HADCRUT4 and GISTEMP. We estimated the slope using a linear regression for the years 1901-2017 for all combinations of the 4 data sets. The uncertainty of the fit for the slope is estimated using the covariance matrix.
The results are shown in the figure and table given below, which will both be included in the manuscript (the table in the appendix). The results indicate that, especially for CNA and ENA, the CMIP5 models overestimate the regional warming compared to observations, as documented in previous articles (e.g. Alter et al., 2017; Donat et al. 2017). Further articles also showed that the CMIP5 models tend to overestimate soil moisture-temperature coupling (Sippel et al. 2017, Vogel et al. 2018), which can lead to an overestimation of projected changes in temperature extremes (Vogel et al. 2018). These biases appear smaller in the newer CMIP6 models (Seneviratne and Hauser 2020). On the other hand, there can be large differences of the observed trend for some of the regions depending on the observational data sets used (e.g. MED and CEU; see Figure 2 below). We discuss the systematic biases in CMIP5 models together with the differences and uncertainties of the observational data sets in the paper.

We are not sure which exact method the reviewer used to estimate observed scaling factors, i.e. which reference he used for global mean temperature, whether linear regression was used, which time periods were considered, and whether he included ocean grid points within the given SREX regions (which we do not for the regional temperatures). Therefore, our results differ from the numbers given in the reviewer comment (see Table 1 below), although they agree on a general overestimation of regional warming per degree of global warming in the CMIP5 models for North America.

Both global temperature data sets (GISTEMP and HadCRUT4) merge near-surface temperatures over land with SSTs over the ocean, which leads to an inconsistency with how global mean temperature is commonly determined for models by taking near surface temperature over ocean and land (as is also done here; see also Cowtan et al., 2015). Therefore, 1 degree of global mean temperature increase in the observations does not correspond to 1 degree from the models (see also IPCC, 2018). We discuss this issue in the paper.

References:

Alter, R. E., Douglas, H. C., Winter, J. M., & Eltahir, E. A. B. (2018). Twentieth century regional climate change during the summer in the central United States attributed to agricultural intensification. *Geophysical Research Letters*, 45, 1586-1594. https://doi.org/10.1002/2017GL075604

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M. E., Miller, S. K., Steinman, B. A., Stolpe, M. B., and Way, R. G. (2015), Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, Geophys. Res. Lett., 42, 6526-6534, https://doi.org/10.1002/2015GL064888.

Donat, M.G., A.J. Pitman, and S.I. Seneviratne (2017). Regional warming of hot extremes accelerated by surface energy fluxes, Geophys. Res. Lett., 44, https://doi.org/10.1002/2017GL073733

IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. P.rtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. P.an, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press. Available from https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_SPM_version_report_LR.pdf

Seneviratne, S.I., and M. Hauser (2020). Regional climate sensitivity of climate extremes in CMIP6 vs CMIP5 multi-model ensembles. Earth's Future, https://doi.org/10.1029/2019EF001474

Sippel, S., J. Zscheischler, M.D. Mahecha, R. Orth, M. Reichstein, M. Vogel and S.I. Seneviratne (2017). Refining multi-model projections of temperature extremes by evaluation against land-atmosphere coupling diagnostics. Earth Syst. Dynam., 8, 387-403, https://doi.org/10.5194/esd-8-387-2017

Vogel, M. M., J. Zscheischler, and S.I. Seneviratne (2018). Varying soil moisture-atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe. Earth Syst. Dynam., 9, 1107-1125, https://doi.org/10.5194/esd-9-1107-2018.

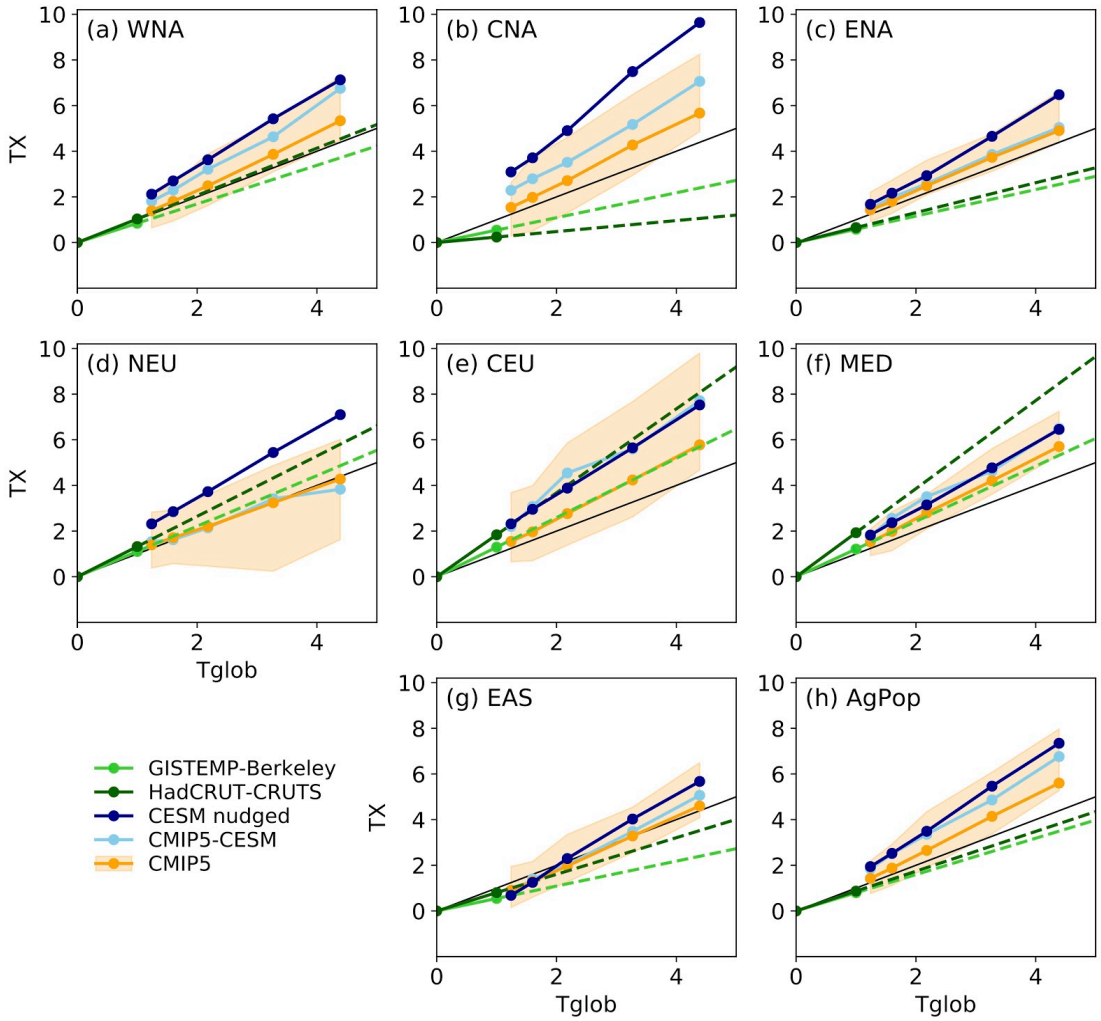

Figure 2: As Fig. 7 in the paper but added observed trends. The solid green lines correspond to the approximate observed warming while dashed green lines indicate the extrapolation beyond the observed warming.

Table 1: The slopes for all combinations of the observational data sets and their uncertainties (one standard deviation).

TXreg	CRUTS		BEST	
Tglob	**GISTEMP**	**HadCRUT**	**GISTEMP**	**HadCRUT**
WNA	0.85 ± 0.21	1.03 ± 0.24	0.85 ± 0.22	1.04 ± 0.25
CNA	0.16 ± 0.30	0.24 ± 0.34	0.55 ± 0.31	0.68 ± 0.35
ENA	0.56 ± 0.22	0.65 ± 0.25	0.58 ± 0.21	0.68 ± 0.23
CEU	1.52 ± 0.30	1.84 ± 0.33	1.30 ± 0.30	1.60 ± 0.34
NEU	1.13 ± 0.31	1.32 ± 0.35	1.11 ± 0.29	1.29 ± 0.33
MED	1.67 ± 0.15	1.93 ± 0.16	1.21 ± 0.16	1.40 ± 0.18
EAS	0.61 ± 0.16	0.80 ± 0.18	0.55 ± 0.19	0.75 ± 0.21
AgPop	0.72 ± 0.12	0.87 ± 0.14	0.80 ± 0.13	0.96 ± 0.14

Minor comments

l.16 The Koreas were also very badly affected.

B2: We included the Korean Peninsula in the list of affected regions.

l.78 How does the end date of July 27 affect the results? Although this captures the largest area with heat, individual regions had heat waves after this date: North Korea experienced its worst heat the first days of August. The Benelux had a second heatwave in early August and the heat on the North American west coast was most severe during the second week of August.

B3: We agree that with the presented study we cannot make statements about heat waves after 27 July 2018, which were more intense in some locations. The choice was also made due to the availability of input files for the atmospheric nudging. Therefore, we cannot provide numbers on how the results are affected by the chosen time period. We changed the text in the results and discussion as well as the conclusions of the manuscript to mention the regions affected by heat waves after July 27 and to discuss this shortcoming of our study.

l.162 I would propose "almost simultaneous", there were weeks differences between these heat waves. Please also mention that there were severe heat waves after the cut-off date.

B4: We agree and use the term “almost simultaneous” as suggested. Also we mention that some regions experienced severe heat waves after the cut-off date (see also answer B3).

l.174 My Newfie friends prefer "Newfoundland".

B5: We thank the reviewer for spotting and correcting this.

l.201 Please mention that in contrast to the CMIP5 model simulations, observed precipitation has increased in CNA over the last century.

B6: We added that the simulated trends are of different sign than the observed summer precipitation trends for CNA.

l.221 Why do you switch from a two-week period to a monthly period? The properties of short-duration heat waves are different from monthly anomalies. Please justify this choice.

B7: For the CMIP5 models the monthly period was chosen for two reasons: First, as the CMIP5 models do not have a prescribed ocean or nudged atmosphere, there is more variability which is better represented using a longer sample and it is also not necessary to have the exact days as every day of July matches conditions of the study time period. Secondly, it is more practicable to process the monthly instead of the daily data. For the nudged experiments in our study we agree that the choice of the time period and its length could have been expected to affect the results, although we find that the effects are actually very small. To assess this point, we repeated the analysis plotting only July 13-27 for CESM nudged. The results show that changing the monthly to a two-week period for the nudged experiments does not substantially impact the results (see Figure 3 below). The change is largest for ENA where a reduction of the slope can be found. However, results are still qualitatively the same.

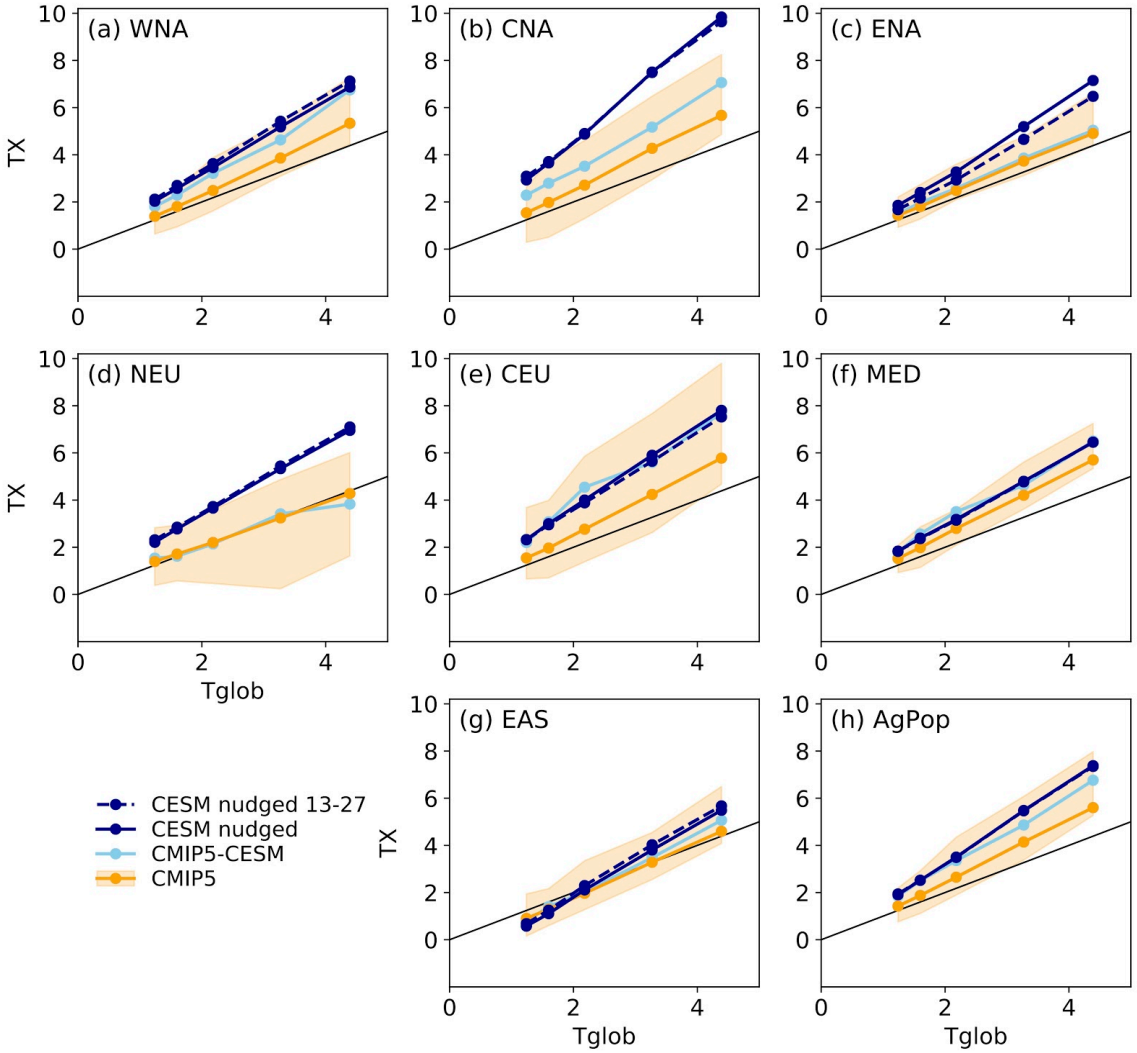

Figure 3: Same as Figure 7 in the paper but showing July 13-27 for the nudged simulations (dark blue; stippled) and July 1-27 (dark blue; solid).

Storylines of the 2018 Northern Hemisphere heat wave at pre-industrial and higher global warming levels

Kathrin Wehrli[1], Mathias Hauser[1], and Sonia I. Seneviratne[1]

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland

Correspondence: Kathrin Wehrli (kathrin.wehrli@env.ethz.ch)

Abstract. Extreme temperatures were experienced over a large part of the Northern Hemisphere during the 2018 boreal summer (hereafter referred to as "NH2018 event"), leading to major impacts to agriculture and society in the affected countries. Previous studies highlighted both the anomalous atmospheric circulation patterns during the event and the background warming due to human greenhouse gas emissions as main drivers for the event. In this study, we present Earth System Model experiments investigating different storylines of the NH2018 event given the same atmospheric circulation and alternative background global warming for: no human imprint, the 2018 conditions, and different mean global warming levels (1.5°C, 2°C, 3°C, and 4°C). The results reveal that the human-induced background warming was a strong contributor to the intensity of the NH2018 event, and that resulting extremes under similar atmospheric circulation conditions at higher levels of global warming would reach very dangerous levels. ~~About 32% (61%~~Compared to 9% during the NH2018 event, about 13% (34%) of the inhabited or agricultural area in the investigated region would reach daily maximum temperatures over 40°C under 2°C (4°C) of global warming and similar atmospheric circulation conditions.

1 Introduction

During the 2018 boreal summer, large parts of the Northern Hemisphere experienced extreme temperatures, leading to major impacts to agriculture and society in the affected countries (Vogel et al., 2019). The event (hereafter referred to as "NH2018 event") was associated with near-simultaneous heat waves on three continents, including North America, Western and Northern Europe, as well as Japan and the Korean Peninsula (Kornhuber et al., 2019b; Vogel et al., 2019). Previous studies highlighted both the anomalous atmospheric circulation patterns during the event and the background global warming due to human greenhouse gas emissions as main drivers for the event (Drouard et al., 2019; Kornhuber et al., 2019b; Toreti et al., 2019; Vogel et al., 2019). The NH2018 event was characterized by a hemisphere-wide wavenumber 7 circulation pattern, which was also observed during the European heat waves of 2003, 2006 and 2015 (Kornhuber et al., 2019b). A strong positive mode of the North Atlantic Oscillation contributed significantly to the extreme summer conditions in Europe by amplifying the weather

anomalies induced by the wavenumber 7 pattern (Drouard et al., 2019). An analysis of simulations from the 5th phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012) showed that the total area affected by hot extremes during the NH2018 event, despite being unprecedented in the historical record, was actually consistent with the present-day level of global warming (Vogel et al., 2019). Indeed, it had approximately a 16% probability of occurrence under present global warming in the CMIP5 simulations (Vogel et al., 2019). However, no studies were conducted so far to disentangle the contribution of the anomalous circulation patterns vs. background global warming for the climate anomalies during the NH2018 event.

In this study, we present numerical experiments investigating different storylines of the NH2018 event given alternative background global warming but the same atmospheric circulation anomalies. While it cannot provide information on probability, the storyline approach allows to explore ~~uncertainty of future climate and~~ the consequences of a specific event ~~in order~~ for different levels of future climate warming to improve understanding of the driving factors involved (Hazeleger et al., 2015; Shepherd et al., 2018). The alternative background global warming conditions applied in the experiments include: a) no human imprint (i.e. natural/ pre-industrial climate conditions), b) 2018 conditions (corresponds to approximately 1.1°C of global warming in the CMIP5 multi-model mean), c) 1.5°C, d) 2°C, e) 3°C and, f) 4°C of global warming. The atmospheric circulation in the experiments is nudged to the observed wind patterns during the NH2018 event following the approach of ~~Wehrli et al. (2019)~~Wehrli et al. (2018). Hence, all of the experiments include the same circulation patterns but different background global warming. These experiments are of particular relevance since events associated with the type of circulation patterns from the NH2018 events could lead to high risks of crop failures across several breadbasket regions of the world (Kornhuber et al., 2019a).

2 Model and methods

Global climate model simulations are conducted with the Community Earth System Model version 1.2 (CESM, Hurrell et al., 2013). Historical sea surface temperatures (SSTs) and sea ice fractions are prescribed using transient monthly observations from a merged product combining the Hadley Centre sea ice and SST data set, version 1 (HadISST1) up to 1981 with the weekly optimum interpolation SST analysis version 2 by the National Oceanic and Atmospheric Administration ~~(NOAA OIv2) thereafter (Hurrell et al., 2008)~~(Hurrell et al., 2008, NOAA OIv2 hereafter). We produce SSTs and sea ice consistent with the different background climates as described in Sect. 2.4 and Sect. 2.5. To simulate the Earth's atmosphere, CESM utilizes the Community Atmosphere Model version 5.3 (CAM5, Neale et al., 2012). Here we couple CAM5 to a nudging module to control the atmospheric circulation as described in Sect. 2.2. For the representation of land surface processes, CESM uses the Community Land Model version 4 (CLM4, Lawrence et al., 2011; Oleson et al., 2010). CAM5 and CLM4 are both run on 0.9° x 1.25° horizontal resolution with 30 layers up to 2 hPa for the atmospheric component (CAM5) and 10 hydrologically-active soil layers down to 3.8 m for the land component (CLM4). Solar forcing follows the model default historical data until the end of 2005 and historical and future simulations of the forcing compiled for CMIP6 thereafter, as in ~~Wehrli et al. (2019) (see Matthes et al., 2017, for solar forcing)~~Wehrli et al. (2019, for CMIP6 solar forcing see Matthes et al., 2017).

Aerosols and land use/ vegetation follow CMIP5 recommendations for the 20th century simulation until the end of 2005 and the ~~representative concentration pathway~~ Representative Concentration Pathway 8.5 thereafter (RCP8.5, van Vuuren et al., 2011). Likewise, greenhouse gases (GHGs) follow CMIP5 historical recommendations and then after ~~2006~~ 2005 they are prescribed from observations for CO_2, CH_4, N_2O, and RCP8.5 for other GHGs. Observed global means of CO_2, CH_4, and N_2O were obtained from NOAA (CO_2: Dlugokencky and Tans, 2019; CH_4: Dlugokencky, 2019; and N_2O: NOAA Earth System Research Laboratory, 2019).

Each experiment is run for four years; ~~the~~ thus for the historical simulation the first three years (2015–2017) are used as spin-up, and 2018 is analysed. A climatology for the historical simulations is obtained from a longer simulation that covers the years 1981–2010.

2.1 Natural and warming scenarios

The NH2018 event serves as a reference to investigate its ~~resulting~~ characteristics in hypothetical conditions, or storylines, with the same atmospheric circulation but different levels of global warming. In addition to the historical simulation we run one simulation with pre-industrial-like conditions ("natural") and four simulations that follow global warming scenarios. An overview is given in Table 1. For the natural simulation volcanic aerosols and solar radiation are set to historical conditions, whereas we use pre-industrial GHGs, aerosols and SSTs (Sect. 2.4). The four warming scenarios are designed to match 1.5°C, 2°C, 3°C and 4°C global mean warming with respect to pre-industrial conditions (1861–1880) of the CMIP5 multi-model-mean (MMM). We will hereafter refer to these experiments as warming15, warming20, warming30 and warming40. Aerosols, GHGs and SSTs follow RCP8.5. The actual warming of the scenarios slightly differs from the target values, which will be discussed in Sect. 4. All simulations are nudged towards 2015–2018 atmospheric circulation (1981–2010 for the climatology).

2.2 Nudging of the atmospheric circulation

To impose the large-scale circulation of the event year (2018) in the model, we use atmospheric nudging of the horizontal winds. The approach is described and validated in Wehrli et al. (2018) and Wehrli et al. (2019). The horizontal winds are relaxed toward observations using a height-dependent nudging function (for the profile see Wehrli et al., 2018). At the surface the nudging strength is set to zero, meaning that the land surface can interact with the atmosphere through surface turbulent fluxes, resulting in balanced surface climate and winds. The large-scale circulation (mostly above 700 hPa) is forced to follow the observations and thus ensures that the observed large-scale weather patterns are reproduced (see Kooperman et al., 2012; Wehrli et al., 2018; Zhang et al., 2014). As a proxy for the observed winds, we use zonal and meridional 6-hourly wind fields from the ERA-Interim reanalysis (Dee et al., 2011). The nudging of the circulation ends on July 27th 2018 due to availability of the input fields.

Table 1. Overview of simulations

Name	Year	Atmospheric forcing	CMIP5 MMM warming [°C]	Actual warming [°C]
climatology	1981–2010	historical+RCP8.5	–	–
historical	2018	historical+RCP8.5	1.12	1.24
natural	2018	pre-industrial+historical	~~0.0~~ 0.00	~~0.00~~0.00
warming15	2028	RCP8.5	~~1.5~~ 1.50	1.60
warming20	2042	RCP8.5	~~2.0~~ 2.00	2.18
warming30	2064	RCP8.5	~~3.0~~ 3.00	3.27
warming40	2085	RCP8.5	~~4.0~~ 4.00	4.39

Year corresponds to the year analysed in this study and atmospheric forcing refers to the solar, aerosol, and greenhouse gas forcing. The warming of the CMIP5 multi-model mean (MMM) is given relative to a pre-industrial time period (1861–1880) and corresponds to the target warming level for the respective simulation years. Per design the temperature for the natural simulation is set as reference for no global warming. The actual warming in the CESM simulations differs from the CMIP5 MMM.

2.3 Determination of warming levels in CMIP5

To produce the ocean fields for the natural and warming scenarios we use model output from the historical and RCP8.5 emissions scenarios from the CMIP5 data archive. We use one simulation per model ("r1i1p1"). To find the years corresponding to the chosen warming scenarios in the CMIP5 ensemble we use near-surface air temperature ("tas"). The pre-industrial reference period is given as 1861–1880. Warming levels with respect to pre-industrial are then determined by taking the difference between annual global mean temperature from RCP8.5 and pre-industrial for each model individually. A 21-year centred running mean is applied to the differences and we then compute the MMM. Following this approach, we obtain a CMIP5 MMM warming of 1.12°C for 2018. The 1.5°C, 2°C, 3°C and 4°C MMM warming levels are reached in 2028, 2042, 2064 ~~,~~ and 2085, respectively.

2.4 Sea surface temperatures representative of prescribed warming levels

To derive SSTs consistent with the different background climates we add delta SST fields to the observed monthly SSTs. The delta SST fields are computed from the CMIP5 SST fields ("tos"), which were regridded to a common grid of 1° x 1°. We first apply a 21-year running mean over the monthly merged historical + RCP8.5 tos data and then average over the models. ~~Using these~~ This way, temporally-smoothed monthly fields ~~between 1979–2020 and for~~ are created for 2018 (as well as for the three year spin-up: 2015–2017) and the years matching the warming levels: 2028, 2042, 2064 and 2085 ~~,~~ (as well as three years spin-up for each). Transient delta SSTs for the warming scenarios are then computed by subtracting the temporally-smoothed fields for the present-day period (2015–2018) from the fields for the warming levels (e.g. 2025–2028 for the 1.5°C warming simulation). A monthly climatology is computed over the ~~monthly climatology of the~~ pre-industrial ~~tos fields we compute the~~

~~transient delta SSTs by taking the differences~~reference period (1861–1880) and the delta SSTs for the natural simulation are then computed by subtracting the present-day period from this pre-industrial climatology. These delta SSTs (see also Fig. 2a) are then added to the ~~NOAA OIv2 historical data~~ historical SSTs of the model to create the SSTs that are prescribed in our scenarios. The detailed step-by-step procedure is given in the Appendix (B1).

2.5 Generation and prescription of sea ice

Although it would be possible to derive a delta field for sea ice similarly to the SSTs, this would result in a sea ice field that is not in balance with the new SSTs. Therefore, we derive a relationship between sea ice fraction anomalies and SST anomalies with respect to a climatology, similarly as in the "Half a degree additional warming, prognosis and projected impacts" experiments (HAPPI; Mitchell et al., 2017). Monthly SST and sea ice anomalies are computed for the years 1996–2015 from the climatology of the same years, using ocean observations from NOAA OIv2. A linear regression is then fitted to the anomalies for each month-of-the-year, longitude and for both hemispheres separately. ~~The slope and intercept are smoothed zonally first, before they are then used to compute sea ice anomalies for the natural and warming scenarios as a function of the delta SST field.~~ As an extension of the method applied in the HAPPI experiments we only consider grid cells that undergo a change of sea ice fraction of over 50% for the month in consideration. This ensures that grid cells that are not experiencing enough variability during 1996–2015 are excluded from the analysis. For example, a grid cell close to the ~~north pole~~ North Pole may always have a sea ice fraction larger than 90% while the ocean temperature usually ~~does not change at all~~ changes only minimally and does therefore not allow for a robust estimation of the slope and intercept of the relationship. Consequently, the given grid cell would not melt even under very high global warming. To compute the regression of a grid cell we pool SST and sea ice anomalies of all valid grid cells that are within three grid cells to the west and to the east and along the ~~meridian~~ meridians in the same hemisphere. Should no valid grid cells be in this area, more grid cells in the longitudinal direction are included gradually (extending box approach). If a maximum of 11 grid cells to the west and to the east is reached and no valid grid cells are found, a hemispheric linear fit is used. The slope and intercept from this regression are smoothed ~~along the longitude~~ zonally using a 500 km smoother as is done for HAPPI. We further tested an approach where not all grid cells along a meridian were included, but up to two grid cells to the north, south, east, and west of a specific grid cell (5 x 5 grid cells in total). The results are very similar to the first method (not shown). We choose the first method because the resulting sea ice field is overall smoother. The new set of SSTs and sea ice for each scenario is adjusted according to the constraint of Hurrell et al. (2008), which ensures that (i) sea ice fraction is 100% at $-1.8°$C, and SSTs do not get colder than that, (ii) there is no sea ice at water temperatures warmer than $4.97°$C, and (iii) that within this temperature range the maximum sea ice fraction is limited by a temperature-dependent function.

We contrast our method to the one developed for the "Climate of the 20th Century Plus Detection and Attribution" project (C20C+DA; Stone et al., 2019). For this method, a linear relationship is determined using absolute values of SST and sea ice coverage instead of anomalies. The regression is calculated by pooling all ice-covered grid cells of the Northern and Southern Hemisphere, respectively. Ice coverage is binned in 100 equally-sized bins and the median SST for each ice bin is determined. The line through the centre of mass of all the bin medians is then estimated by a linear fit. This method was developed to

compute sea ice estimates for natural historical simulations (i.e. a cooling). Therefore, it comes with an algorithm that prevents ice from melting and only adds new ice where SSTs cool. As we require a method that works for positive and negative delta SSTs we do not implement the full algorithm but only make use of the linear relationship. For ~~20C~~C20C+DA the years 2001–2010 were used to determine the relationship. For consistency we take the years 1996–2015. Again, we apply the constraint of Hurrell et al. (2008). We hereafter refer to this ~~method as SP equation~~as the C20C method.

2.6 Data sets and data analysis

The atmospheric nudging uses ERA-Interim 6-hourly horizontal wind fields. Mean daily near-surface temperature fields are retrieved from ERA-Interim for comparison to our results. Observed daily maximum near-surface temperatures (TX) are obtained from ERA-Interim and Berkeley Earth (Rohde et al., 2013a, b). Mean daily near-surface temperature, TX and precipitation are retrieved from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2, Gelaro et al., 2017). The ~~observations/ reanalysis data is~~ data sets are remapped to the resolution of CESM using second-order conservative remapping (Jones, 1999). Results are shown for the absolute values of the variables as well as for anomalies with respect to the 1981–2010 climatology.

We estimate observed regional trends in TX as a function of global mean temperature. Therefore, we use TX from Berkeley Earth and the Climate Research Unit (CRU) data set CRU TSv4.03 (Harris et al., 2020). For global mean temperature GISTEMPv4 (GISTEMP Team, 2020; Lenssen et al., 2019) and HadCRUT4 (Morice et al., 2012) are used. The regional averages are computed using the original resolution of CRU TSv4.03 and Berkeley Earth. The regional trends are computed using a linear regression for 1901–2017. Uncertainties of the trend are estimated using the covariance of the residuals of the fit. This gives four trend estimates, whereof only two are shown later on, corresponding to the steepest and flattest estimates for most regions. All possible estimates and their uncertainties are given in Table A1.

2.7 Bias correction

Absolute values for TX are bias-corrected using empirical quantile mapping, which corrects the entire distribution of a variable as described in Déqué (2007). We apply the implementation by Rajczak et al. (2016), which is available in the R package qmCH2018 (see also "Code availability"). We choose 1981–2015 as the calibration period and the quantile mapping correction is calibrated for each day-of-the year using a 91-day moving window. The model distribution is translated to the distribution of a reference using 99 quantiles. For model values that are more extreme, quantile mapping uses the correction function of the 99th (1st) quantile to correct values above (below) the calibrated quantiles (Themeßl et al., 2012). We used Berkeley Earth and ERA-Interim as reference for the modeled TX. In the following results for Berkeley Earth are shown but the results for ERA-Interim are very similar (differing by less than 2.5% for the regional averages of the study region).

A mean bias correction specific to the day-of-year was tested but results were found to agree less with the reference data from ERA-Interim and Berkeley Earth (not shown). Note that only absolute TX values are bias-corrected. In cases where anomalies are shown for TX as well as other variables, no correction of the original model output is performed.

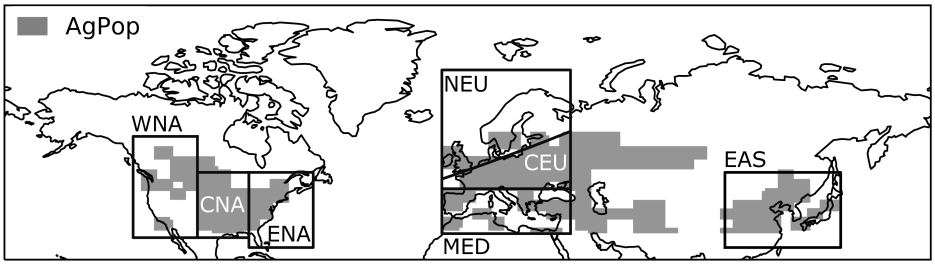

Figure 1. Study regions. Shown by the grey shading are regions north of 30°N with high population density and/ or high importance for agriculture (AgPop). The black outlines mark the location of selected SREX regions in America ~~and~~ , Europe and Eastern Asia. Note that the Eastern Asia region (EAS) was cropped and extends only from 25°N to 50°N.

2.8 Study regions

We show results only north of 25°N. For regional averages we choose the regions defined in the IPCC Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX; Seneviratne et al., 2012). We show results for the European and American regions as well as the Eastern Asia region (EAS) as highlighted in Fig. 1. ~~In addition, we~~ Note that the southern extent of EAS was cropped such that it only extends from 25°N to 50°N. Ocean grid points are excluded from the analysis. We also focus on a region north of 30°N that is especially vulnerable to extreme conditions because it is either densely populated (more than 30 km^{-2}) and/ or an important area for agriculture. We define this "human-affected and human-affecting" region following Seneviratne et al. (2018) and Vogel et al. (2019) and refer to it as AgPop (see Fig. 1).

3 Results and discussion

We first evaluate the sea ice reconstruction method introduced in this study against observed sea ice and compare the performance to the ~~SP equation~~ C20C method in Sect. 3.1. Then we present the results for the natural and warming scenarios in Sect. 3.2 and discuss their implications for possible future events. In Sect. 3.3 we discuss how TX scales with an increase in global mean temperature and if the NH2018 atmospheric circulation influences this relationship.

3.1 Evaluation of sea ice reconstruction

Historical sea ice fields are ~~computed~~ modelled using the methods described in Sect. 2.5 and evaluated against NOAA OIv2 ice (see Appendix Fig. A1). While the ~~method using the SP equation~~ C20C method generally overestimates sea ice, over- and underestimation nearly balance out when taking a time average over multiple years for the method presented here ~~.~~ (Fig. A1a,c). Hence, the spatial root mean squared error (RSME) for ice grid cells is $< 1\%$ using the new method while it is around 5% using the ~~SP equation~~C20C method. For a single year the errors are larger, ~~5.6% and~~ 8.5% and 5.6% for the year 2018 for the ~~new method and SP, respectively .~~ C20C and new method, respectively (Fig. A1b,d). It is not possible to evaluate the performance

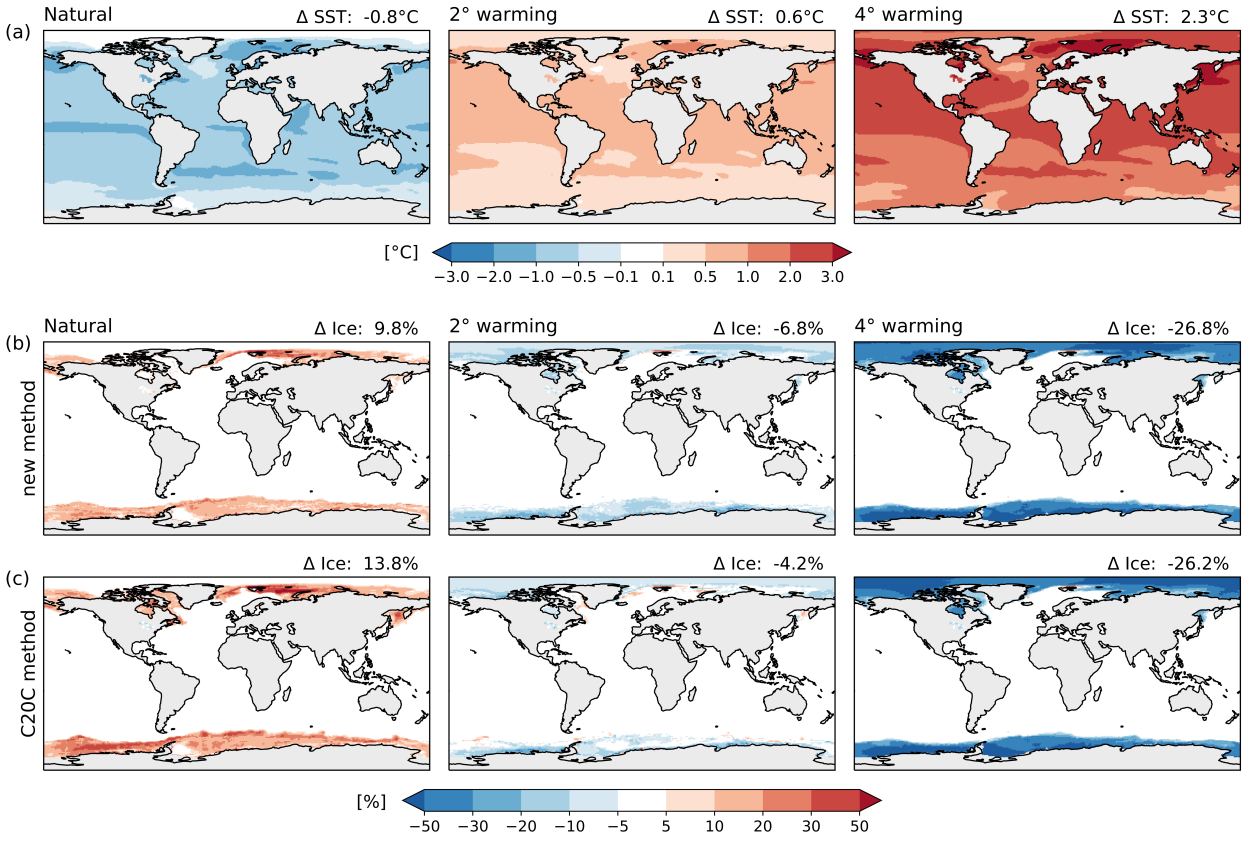

Figure 2. Change in SST and sea ice coverage with respect to historical conditions from NOAA OIv2 for the natural (left), warming20 (middle) and warming40 (right) scenarios. (~~top row~~a) Average delta SST for 2018 under the three climate scenarios. (~~centre row~~b) Average sea ice coverage change for 2018 under the three climate scenarios using the sea ice estimation method developed in this study. (~~bottom row~~c) ~~same~~ Same as in the ~~centre row~~ (b) but using the ~~SP equation~~ C20C method to compute the change.

of the sea ice reconstruction for the natural and warming scenarios. The two methods suggest a change of $+10\%$ to $+14\%$ in the annual mean with respect to currently observed sea ice fraction for the natural scenario, -4% to -7% for the 2°C warming scenario and -26% to -27% for 4°C warming (Fig. 2~~,~~ b,c, see Appendix Fig. A2 for the absolute ice fields for the new method). Note that the two methods show larger differences for the natural scenario but agree ~~more~~ better for the warming scenarios.

3.2 Storylines for the NH2018 event

During late June and the first weeks of July 2018 an exceptionally strong and persistent Rossby wavenumber 7 pattern dominated the weather in the Northern Hemisphere (Drouard et al., 2019; Kornhuber et al., 2019b). The wave pattern was associated

with almost simultaneous extreme events across the entire hemisphere including heat waves in the western United States, eastern Canada, ~~western~~ Asia and large areas in Europe (Kornhuber et al., 2019b; NOAA, 2018; Vogel et al., 2019). In Europe, the July temperature anomaly was ranked second highest on record, just one hundredth degree Celsius behind 2015 (NOAA, 2018). Vogel et al. (2019) show that the area of important "human-affected and human-affecting" regions (AgPop) experiencing simultaneous heat waves peaked at the end of July. Hence, we focus here on the month of July 2018 and especially on a 15-day period at the end of the month from July 13 to 27 (which also corresponds to the last days of our simulations where the atmospheric nudging is available). Note that strong heat waves also occurred after July 27, which were locally more extreme than the heat waves examined during our study period. For example in Spain and Portugal the heat wave peaked during the first week of August (Barriopedro et al., 2020), the Netherlands experienced a second heat wave starting at the end of July and lasting for the first week of August (KNMI, 2020), and South Korea reported new record daily maximum temperatures in the beginning of August (KMA, 2019).

Comparison of the temperature anomaly to ERA-Interim shows that the daily and 15-day mean anomalies are well represented in the nudged historical CESM simulation (Fig. 3). Similar results are obtained for MERRA-2 (not shown). For 13 to 27 July 2018 the mean daily temperature in the historical simulation is on average 5°C warmer than the climatology (1981–2010) for large areas in Scandinavia and some smaller areas in Germany, Belgium and the Netherlands (~~maps in~~ Fig. 3i). In Northern America temperatures are ~~2.5~~ 2.5°C to 5°C above average for ~~Neufundland~~Newfoundland, Québec, Texas and northern Mexico as well as ~~most~~ large parts of the Western North American (WNA) ~~region~~and EAS regions. An intensification of the hot anomaly can be seen for the Central European (CEU) and Northern European (NEU) regions during July ~~,~~(Fig. 3d,e), whereas in the AgPop region the anomaly is around 1.5 to 2°C for the entire month (~~time series plots in~~ Fig. 3h). TX over the same time period is simulated even more than 7.5°C warmer than the climatology in much of Scandinavia and Northern Germany, Belgium and the Netherlands (Fig. 4b and Appendix Fig. A3a-d for comparison to ERA-Interim, MERRA-2 and Berkeley). In contrast, TX for much of the Mediterranean (MED) region is close to the climatological average. Colder-than-average TX is seen for areas surrounding the Aegean Sea and Black Sea, as well as for Portugal, parts of Spain and for the United States east coast (Fig. 4~~).~~ b). Note that in Portugal and Spain a heat wave developed in the first week of August (Barriopedro et al., 2020), which is, however, outside our study period. The temporal evolution of the TX anomaly during the month of July resembles that of the daily mean temperature for the study regions (Appendix Fig. A4).

In the nudged historical simulation maximum daytime temperatures (TXx) exceeding 40°C are simulated for parts of the Central North American (CNA) ~~region and for a few locations in northern France, Belgium and the Netherlands~~and WNA regions and east of the Caspian Sea. The fraction of the AgPop region affected by TXx > 40°C is ~~20%. In the natural scenario mean TX anomalies exceed 7.5°C virtually nowhere (except for five grid cells) in the entire Northern Hemisphere north of 25°N (~~9% (Fig. 4~~). TXx during 13 to 27 July does not exceed 40°C for the European regions and only 12%~~ a). In the natural scenario only 7% of the AgPop region is affected by such high temperatures. ~~Although the circulation pattern induces anomalous temperatures in the same regions as in the historical case, very extreme temperatures > 40°C are confined to areas that climatologically have higher likelihood to experience hot temperatures such as northern Africa.~~ On the other hand, the warming scenarios show ~~an extreme~~ a strong intensification of the magnitude and extent of the event. In the warming20

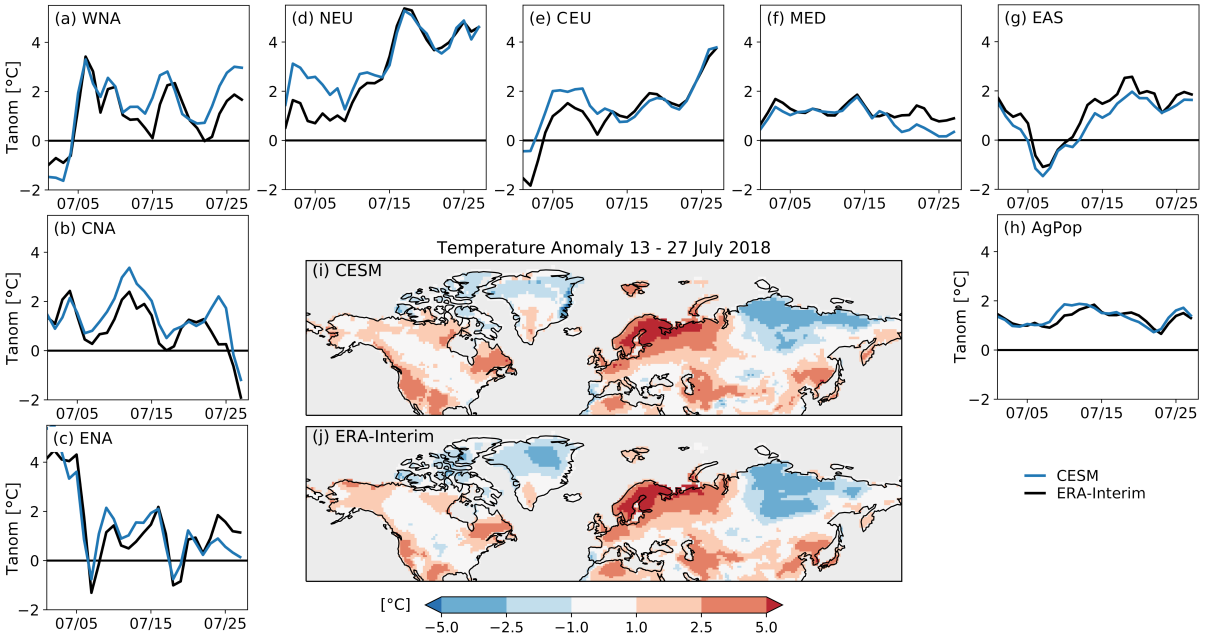

Figure 3. Anomalies of ~~mean~~ daily mean temperature with respect to the 1981–2010 climatology. The maps show the average anomaly over 13 to 27 July 2018 for (i) the nudged historical CESM simulation and (~~top~~j) ~~and~~ for ERA-Interim~~(bottom)~~. The time series show the evolution of daily temperature anomaly (Tanom) for CESM (blue) and ERA-Interim (~~grey~~black) for the month of July averaged over ~~six~~ (a-g) seven SREX regions as well as (h) for the AgPop region.

scenario the fraction of the AgPop region experiencing > 40°C temperatures increases to ~~nearly one third~~ 13% of the total area and ~~it almost doubles to 61% for~~ more than one third (34%) is affected in the warming40 scenario. At 4°C global warming the model predicts that most of the United States will experience temperatures above > 40°C given the 2018 circulation pattern (Fig. 4a). For much of CNA anomalies with respect to 1981–2010 exceed 10°C ~~.~~ (Fig. 4b). Large areas of CEU and NEU experience temperatures around 38°C and higher, which corresponds to more than 10°C above climatology for Scandinavia.

The NH2018 event did not only bring exceptionally warm temperatures to central and north-western Europe but also a dry spring and summer contributed to severe drought conditions (Toreti et al., 2019). In contrast, south-eastern Europe experienced a wetter-than-usual spring and summer (Toreti et al., 2019). This marked precipitation dipole over Europe is reproduced in the historical simulation (Fig. 5). In ~~addition, over North America a contrasting precipitation pattern can be found with a strong positive anomaly in the east and a precipitation deficit in the west (Fig. 5). In the~~ the warming40 scenario we find a decrease of up to −60% in precipitation over Mediterranean Europe ~~and CNA~~ compared to the historical simulation (Fig. 6~~, left). For Europe these~~ a). These changes in precipitation counteract the precipitation dipole observed in the historical simulations but precipitation still remains above climatology for most of Mediterranean Europe ~~. Over a~~ (not shown). Additionally, the warming scenarios simulate higher net surface shortwave radiation (Fig. 6b) in large parts of Europe resulting in an increase in sensible heat flux and a decrease in latent heat flux (not shown). In Eastern Asia very dry conditions are observed and simulated for

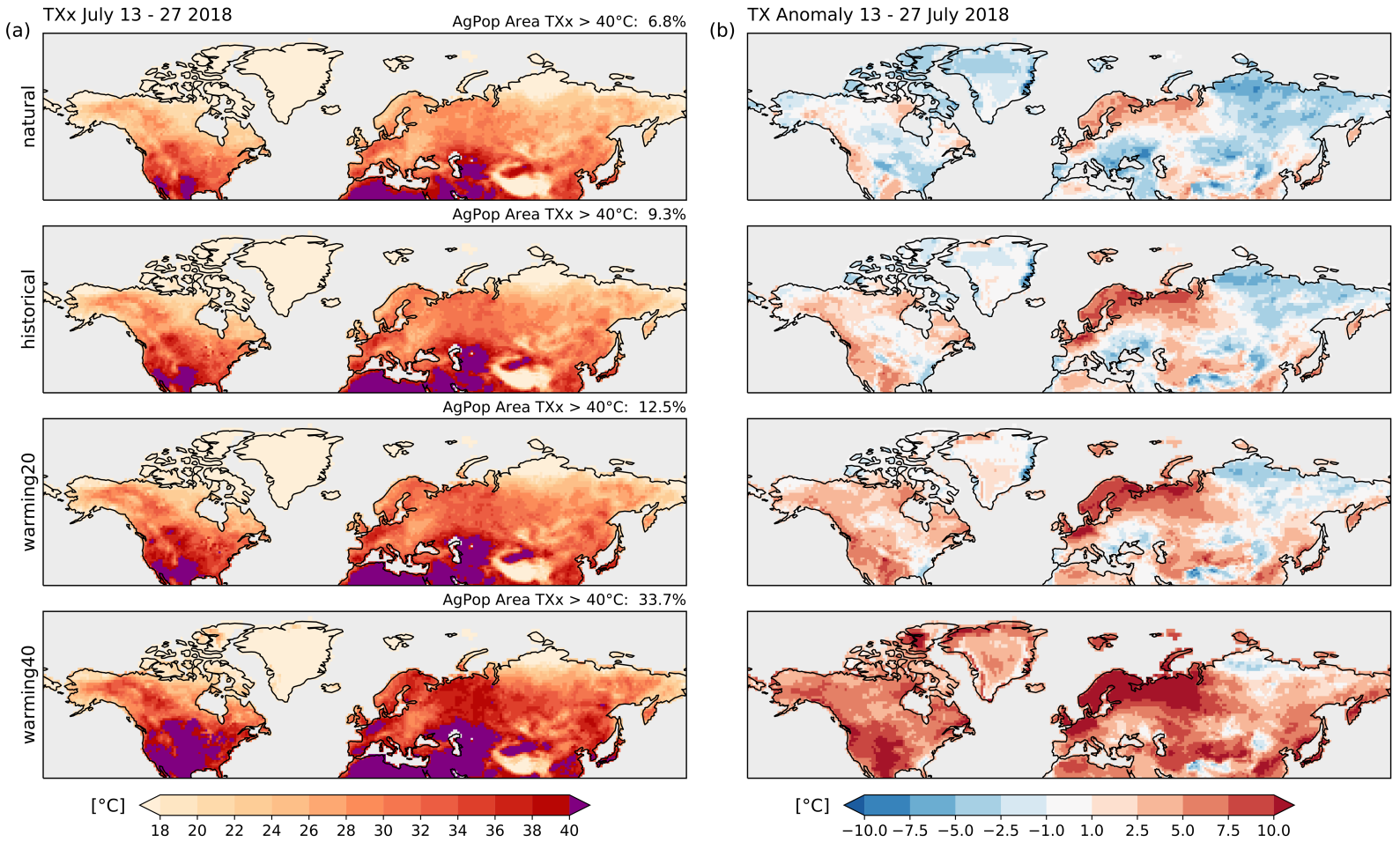

Figure 4. Daily maximum ~~temperature~~ temperatures (TX) for July 13 to 27 2018. (~~left~~a) ~~Maximum~~ Bias corrected absolute maximum TX (TXx) over the 15-day period. The ~~model simulations were corrected by the mean (1981–2010) climatological TX bias compared to Berkeley Earth. The~~ number in the upper right corner indicates the fraction of the AgPop ~~area~~ region where absolute temperatures exceed 40°C. (~~right~~b) Anomaly of mean TX over the 15-day period compared to the 1981–2010 historical climatology (~~uncorrected~~no bias correction).

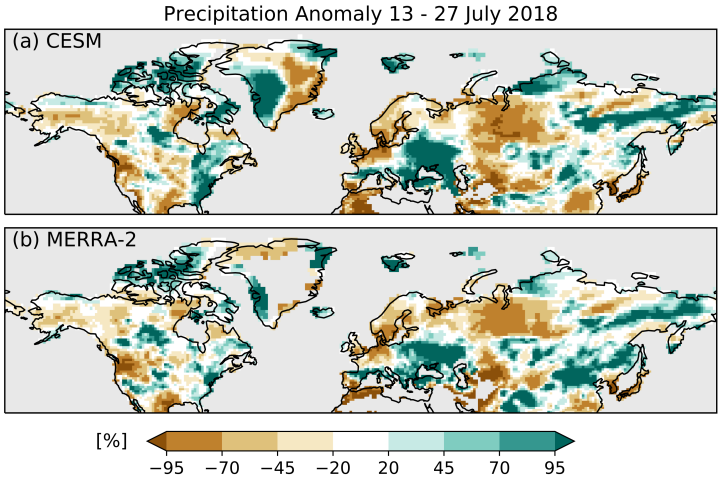

Figure 5. Precipitation ~~Anomaly~~ anomaly (%) for July 13 to 27 2018 compared to the 1981–2010 climatology. (~~top~~a) CESM historical simulation (~~bottom~~b) MERRA-2. ~~Areas~~ Land areas with less than 0.1mm/day precipitation in the climatology are masked out.

Japan and the Korean Peninsula, whereas rather more than climatological precipitation can be seen for the neighbouring regions in China (Fig. 5). No substantial precipitation decrease is simulated for the warming scenarios (Fig. 6a) although total cloud cover fraction slightly decreases (not shown). Higher net shortwave radiation (Fig. 6b) leads to an increase in sensible heat flux in the warming40 scenario, especially over Japan and Korea (not shown). Over North America a contrasting precipitation pattern is simulated with a strong positive anomaly in the east and a precipitation deficit in the west (Fig. 5a). Over a smaller region in CNA there is ~~already~~ a precipitation decrease of around ~~40%~~ −40% in the warming20 scenario ~~.~~ and up to −60% in the warming40 scenario. This is in contrast to observed summer precipitation trends showing an increase in the central United States over the last century (Alter et al., 2018; Wuebbles et al., 2017). The precipitation decrease co-occurs with a decrease of up to 25% in total cloud cover fraction for central North America in the warming40 scenario (not shown) as well as higher net shortwave radiation at the surface for large parts of North America ~~and Europe~~ (Fig. 6~~, right~~b). Further, the increase in net shortwave radiation goes along with a decrease in latent heat flux and an increase in sensible heat flux, which is most pronounced in the warming40 scenario for CNA ~~and central Europe~~ (not shown). This change in the surface fluxes implies a reduction in evaporative cooling and increase of near-surface heating, which can amplify the heat wave (Fischer et al., 2007; Seneviratne et al., 2006, 2010). Hence, in a warmer climate the heat ~~wave~~ waves during a NH2018-like event might be amplified through land-surface feedback. Further, some of the regions experiencing precipitation excess might be seeing less precipitation, for example the Mediterranean. This agrees with findings by Toreti et al. (2019) showing that the projected likelihood of anomalously wet conditions as observed during NH2018 decreases for southern Europe. On the other hand, in the natural scenario the event is less extreme due to higher total cloud cover (not shown), less net shortwave radiation (Fig. 6~~, right~~b), an increase of precipitation, especially for CNA (Fig. 6~~, left~~a), and higher soil moisture (not shown). Together with the colder background climate these effects combine to reduce the maximum reached temperatures and abate the heat stress for

example for crops. In short, we find that the NH2018 event would have been less widespread and less hot under natural climate conditions. Contrarily, it would have affected an even much larger area and would have caused particularly dangerously high temperatures and severe drought conditions in a large fraction of the AgPop region under higher levels of global warming.

3.3 Scaling of local temperature increase with global warming

We scale the increase in ~~July mean TX~~ mean TX for 13–27 July of each study region with the global mean temperature change in our scenarios. As our simulations are all nudged toward the 2018 atmospheric circulation, we compare the results to simulations for July mean TX with random circulation from CMIP5 and in specific to the CESM simulations in CMIP5 (CMIP5-CESM). This way it is possible to disentangle the effect of the specific circulation pattern from the global mean warming trend qualitatively (note, however, that the CMIP5 models including CMIP5-CESM ~~uses~~ use an interactive ocean). The simulation years from the CMIP5 models are chosen to equal the actual warming for the CESM simulations (Table 1). In general, the increase in TX with global mean warming between 1.2°C and 4.4°C follows a linear relationship (Fig. 7). For MED both CESM simulations as well as the CMIP5 MMM lie close together and there is little spread between the CMIP5 models (Fig. 7f, orange shading). This indicates that the CESM model behaves similarly to the CMIP5 MMM and that there is no change in the relationship induced by the atmospheric circulation of 2018. Hence, the increase in TX is driven by the background global warming. For WNA and CEU both CESM model configurations simulate a higher TX for a given warming level than the CMIP5 MMM ~~,~~ (Fig. 7a,e), indicating that CESM produces warmer future climate in general for these regions. However, the nudged CESM simulations show a very similar increase of TX as CMIP5-CESM, which implies that for these regional averages the 2018 circulation did not alter the relationship to the global mean temperature. Both SREX regions cover a rather large and climatologically diverse area. In the nudged historical simulation positive TX anomalies are observed in the western part of the CEU region and negative anomalies in the east (Fig. 4b). The two circulation-related contributions likely compensate each other when computing the regional average for the historical simulation and also for the warming scenarios, which might be the reason why the scaling is similar to the non-nudged simulations. For EAS nudged CESM and CMIP5-CESM show higher TX at higher global warming levels than the CMIP5 MMM but not at lower global warming levels (Fig. 7g). The slope of the nudged CESM is steeper than of CMIP5-CESM, indicating a contribution by the atmospheric circulation pattern of 2018. For NEU, CNA and Eastern North America (ENA) TX from the nudged CESM ~~model shows different behaviour from~~ simulations shows a steeper increase with global warming compared to CMIP5 and CMIP5-CESM ~~,~~ (Fig. 7d,b,c), whereas the latter two behave ~~rather similarly.~~ similarly. At the highest level of global warming (4.39°C) TX from the nudged simulations ~~shows a steeper increase with global warming compared to the CMIP5 models and~~ even exceeds the envelope of CMIP5 models~~at the highest level of global warming (4.39°C)~~. Hence, there is an effect of the circulation pattern in these regions which linearly increases for the same event at higher global warming levels. For ENA TX is mostly around climatological values or even below in the historical simulation. Therefore, it is counter-intuitive that for the warming scenarios TX increases due to the same circulation pattern. We speculate that this is related to the increase in shortwave radiation seen for ENA at higher global warming scenarios ~~such as warming40~~ (Fig. 6b). For CNA there is a steeper increase of TX between the 2.18°C and 3.27°C global mean warming which might be related to a change in surface heat fluxes and

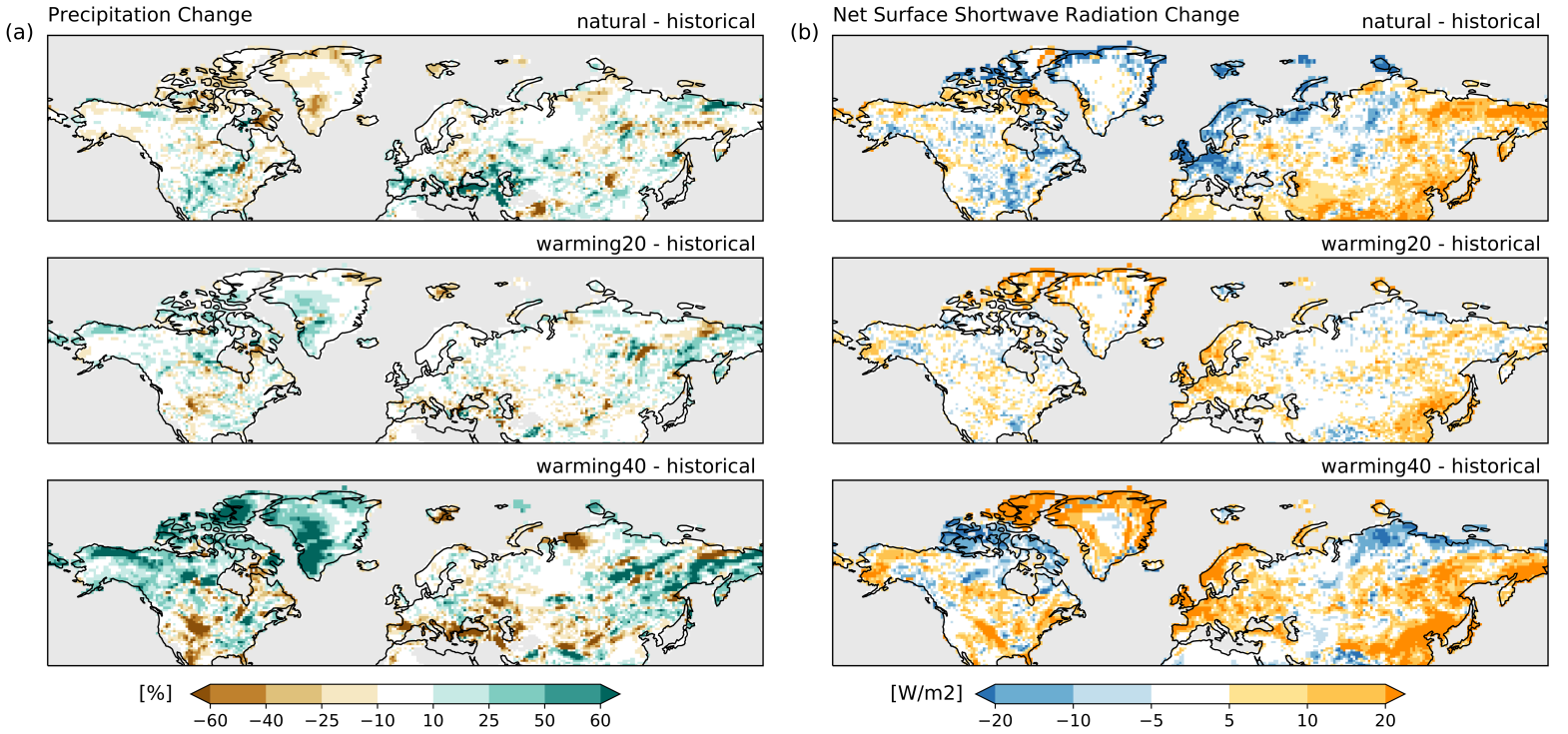

Figure 6. (a) Changes in precipitation (~~left,~~ %); and (b) net shortwave radiation at the surface (~~right,~~ W~~/m2~~ m^{-2}). Changes are averaged over ~~13–27~~ 13 to 27 July 2018 and are given as anomalies from the historical simulation. ~~Areas~~ Land areas with less than 0.1 mm/day precipitation in the historical simulation are masked out in ~~the left plot~~(a).

stronger land-atmosphere coupling as was found in Sect. 3.2. The AgPop region spans across several of the SREX regions evaluated here, mainly EAS, CNA, CEU and MED. Therefore, the relationship is a combination of the effects described above. For the relationship between ~~July mean temperature~~ mean daily temperature in July and global warming the results look similar except for the ~~AgPop region~~ EAS and AgPop regions where the differences between the nudged CESM and CMIP5-CESM disappear (not shown).

We compare the simulated trends to the observed trends over the last century (extrapolated to higher global warming levels). For North America, especially CNA (Fig. 7b) and ENA (Fig. 7c), the regional warming is stronger in the CMIP5 models compared to observations. It has been documented in previous articles that temperature trends are commonly overestimated by CMIP5 models for North America (e.g. Alter et al., 2018; Donat et al., 2017), which was shown to be due to an overestimation of soil moisture-temperature coupling (Sippel et al., 2017; Vogel et al., 2018) and an increase of irrigation that is not included in the models (Alter et al., 2018; Thiery et al., 2020). Warming seems to be overestimated for the same reasons in EAS (Fig. 7g, see also Do In the latest CMIP6 ensemble, these biases appear smaller (Seneviratne and Hauser, 2020). Depending on the observational data sets used, there can be large differences in the observed trend for some of the regions (e.g. CEU and MED: Fig.7e,f). The trend estimates from the four possible combinations of observational data for global mean temperature and TX – including the uncertainty of the fit of the linear regression – are given in Table A1. Observational uncertainty can substantially affect the trend estimate for example due to incomplete coverage and data infilling (Cowtan and Way, 2014); different data quality control and bias correction methodologies (Morice et al., 2012); as well as due to the inclusion and weighting of different observational data and data types. In addition, both global temperature data sets (HadCRUT4 and GISTEMPv4) blend near-surface temperatures over land with SSTs over the ocean, which differs from the procedure for models where near-surface temperatures are averaged over land and ocean (e.g. Cowtan et al., 2015). Hence, models and observations show different global mean warming rates and it was found that it the rate is slower for observations due to the blending of SSTs and near-surface land temperatures, as well as due to very sparse observations of the warming polar regions (about 1.1°C since pre-industrial vs. 1.0°C; see also Cowtan et al., 2015; Rich Therefore, the observed trend shown in Fig. 7 would be smaller if computed in the same way as for the models.

4 Conclusions

We present an analysis of scenario storylines building on the extreme 2018 Northern Hemisphere summer ("NH2018 event"). These storylines retell the NH2018 event in alternative worlds with the same atmospheric circulation as observed but different background global mean warming~~(and associated sea ice cover)~~. The event is alternatively simulated in a natural setting without human imprint on the climate system ("natural"), for the present-day climate conditions, and for four scenarios at different levels of global warming (1.5°C, 2°C, 3°C, and 4°C). All simulations nudge the large-scale atmospheric circulation toward the 2018 conditions but differ in their greenhouse gas and aerosol forcing as well as in the SSTs and sea ice coverage~~of the ocean. These~~ . The focus of this study is on a period from 13–27 July 2018, when the heat wave affected a large fraction of the populated area of the Northern Hemisphere. It has to be noted, that locally severe heat waves were observed during other time periods, mainly in the beginning of August as for example in Korea and for the Iberian Peninsula

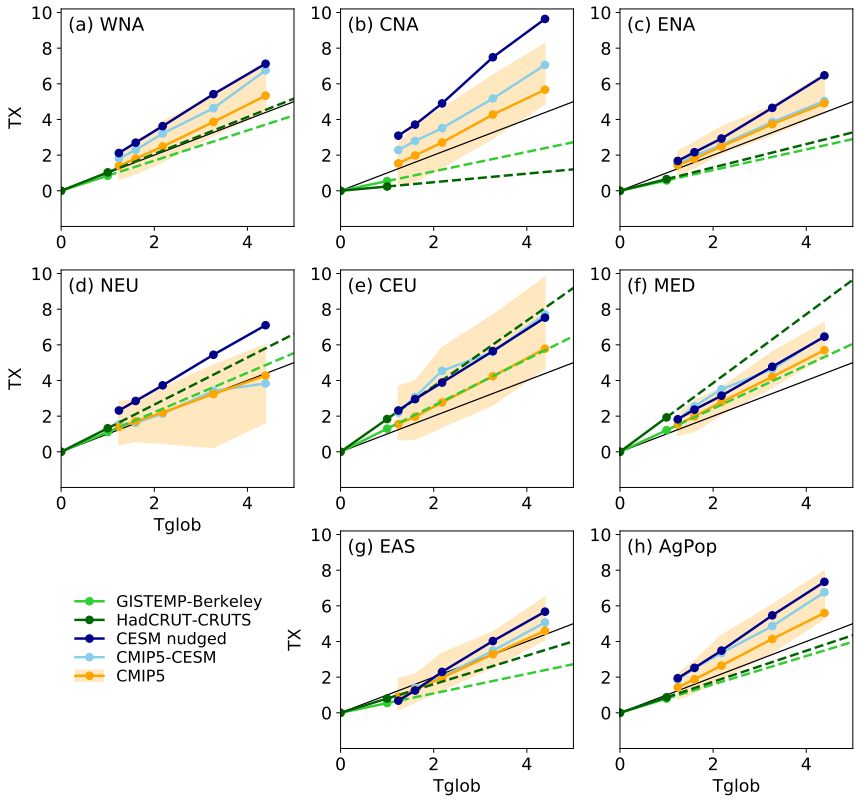

Figure 7. Scaling of daily maximum temperature (TX) for July with global mean warming (Tglob) for the study regions. Shown is the ~~July~~ mean ~~temperature~~ TX for 13–27 July from the nudged CESM simulations (dark blue~~; until July 27~~), the July mean TX for the CMIP5 multi-model-mean (orange line) and the full ~~model~~ range (orange shading) as well as the July mean TX for CESM from the CMIP5 ensemble (CMIP5-CESM, light blue) separately. The solid green lines correspond to the observed warming while the dashed green lines indicate the extrapolation beyond the observed warming (light green for Berkeley Earth using GISTEMPv4 as reference for global mean temperature change and dark green for CRU TSv4.03 using global mean temperature anomalies from HadCRUT4). The black line indicates the 1:1 line.

(e.g. Barriopedro et al., 2020; KMA, 2019). However, due to the availability of input fields for atmospheric nudging, these later events are not analysed. It is likely that similar effects to those described in this study could be found for other regions, when choosing a time period matching the peak of the heat wave. For these more local features of the NH2018 event a regional climate modeling approach might shed more light on the characteristics of future events and would complement the findings of this study.

The storylines for the NH2018 event show drastic consequences for the entire Northern Hemisphere in case of a re-occurrence of this atmospheric pattern at higher global warming. Maximum temperatures increasingly surpass 40°C with large parts of the southern United States experiencing such extreme temperatures already at 2°C global warming. At 4°C almost the entire United States as well as ~~large areas of~~ regions in Western Europe and Eastern Asia are affected~~by such extreme temperatures~~. The total area of important "human-affecting and human-affected" regions (Seneviratne et al., 2018; Vogel et al., 2019) in the Northern Hemisphere (north of 30°N) experiencing temperatures higher than 40°C increases from ~~20%~~ 9% during the NH2018 event to ~~32% and 61%~~ 13% and 34% at a global warming of 2°C and 4°C, respectively. In the "natural" simulation, the fraction of the area affected reduces to 7%. It has to be noted that these values are sensitive to the bias correction method and the reference data set chosen for the calibration of the correction. A quantile mapping method was chosen because it agrees well with observed TX during the heat wave but its validity for the warming (and natural) storylines cannot be tested.

We find that TX for the different scenarios ~~linearly increases~~ increases linearly with global mean warming. For the CNA, NEU and ENA region (and less strong also for EAS) we find a steeper slope of the relationship given the 2018 atmospheric circulation conditions, indicating that these regions are affected by an amplification of the heat wave in a warming climate. ~~However, it~~ Comparison with trends from observations indicates that the regional warming trend might be overestimated in the models, especially for CNA and ENA. Nevertheless, the trends agree well for the other regions and it has to be noted that the observed trends also have to be evaluated with caution, due to uncertainty of the trend fit, observational uncertainty (e.g. Cowtan and Way, 2014; Morice et al., 2012) and due to a different methodology to compute global mean temperatures in observations compared to models (Cowtan et al., 2015). Further, it is important to keep in mind that even if the increase in temperature is linear, the associated impacts would likely not be. Human well-being, crop yields and fire risk for example are related to certain temperature ranges and effects of the heat wave might be amplified once certain thresholds are surpassed. ~~In the "natural" simulation, temperatures remain below 40°C for most regions north of 30°N and stress on plants is further reduced by higher precipitation and soil moisture in the natural climate conditions.~~

It is intrinsic to this kind of simulations that there is no atmospheric variability among ensemble members, which prevents an assessment of the probability of the scenarios. The study is thus not designed to answer the question of how probable it is for a NH2018-like event to re-occur at a certain warming level. Statistically the probability that exactly the same circulation with the same history and evolution during the heat wave will re-occur is small. Further, the nudging approach does not ensure that the circulation pattern is physically in balance with the scenarios for higher global warming or natural conditions. Hence, it might be that it is unlikely that the atmospheric circulation patterns associated with NH2018 event could establish in a warmer climate~~, which we cannot assess in this kind of study~~. However, it has been shown from observations that the occurrence of the driving NH2018 atmospheric circulation pattern, a stationary wavenumber 7 Rossby wave, has increased significantly

in recent years, a possible consequence of the enhanced land-ocean temperature contrast due to global warming (Kornhuber et al., 2019b). Even if there should be no trend in amplitude or persistence of these wave events, associated heat waves in a future climate will be amplified by ~~the~~ global warming (Kornhuber et al., 2019a). From a dynamical perspective, it appears thus probable that similar wave events will occur in a warmer climate and thus the study of such hypothetical events is highly relevant.

The warming levels and corresponding atmospheric forcing were chosen based on the global mean warming using near-surface air temperature of the CMIP5 MMM and from this corresponding delta SSTs were computed. However, this does not imply that the same warming levels will be reached in our model. Indeed, the global mean warming in the simulations is higher. It was shown that computing blended global mean temperature from near-surface air temperature and SSTs together with accounting for incomplete data from observations leads to approximately $0.2°C$ less global warming since the 19th century (Cowtan et al., 2015; Richardson et al., 2016). Applying this to our simulations, the discrepancy between the target warming level and actual warming is reduced. Still, to simulate the target warming with better accuracy a ~~workaround~~ solution could be to choose the years corresponding to a target warming level from CESM itself (using blended global mean temperatures) and also computing the delta SSTs only from CESM (by using the CESM model data from CMIP5 or by running a long, non-nudged simulation with prescribed ocean following e.g. the RCP8.5 scenario). We argue that even if the target warming might be simulated with better accuracy, the results would not differ fundamentally from those found here as we show that local temperatures in the simulations scale approximately linearly with global warming.

To our knowledge this is the first study to use storylines of different warming levels for a specific event in a global climate model setup. Kornhuber et al. (2019a) show that in the future the wavenumber 7 circulation pattern can lead to major risks in breadbasket regions that are important for crop production. The storyline approach presented here provides insightful results that help understand the risks and consequences of similar events in a future climate. Our results highlight that large areas of the Northern Hemisphere will suffer from major heat stress given the same circulation at higher background warming levels, which can have dangerous consequences for agriculture, ecosystems, the economy and also human health.

Code and data availability. The python code for the sea ice algorithm developed in this study is available upon request. For the quantile mapping we used the R package qmCH2018 version 1.0.1 (https://github.com/SvenKotlarski/qmCH2018; Rajczak et al., 2016), which was run in R version 3.3.2 (https://www.r-project.org/). CMIP5 data is available from the the Earth System Grid Federation (ESGF).

Appendix A: Additional figures and table

This section provides additional figures and tables accompanying the main article.

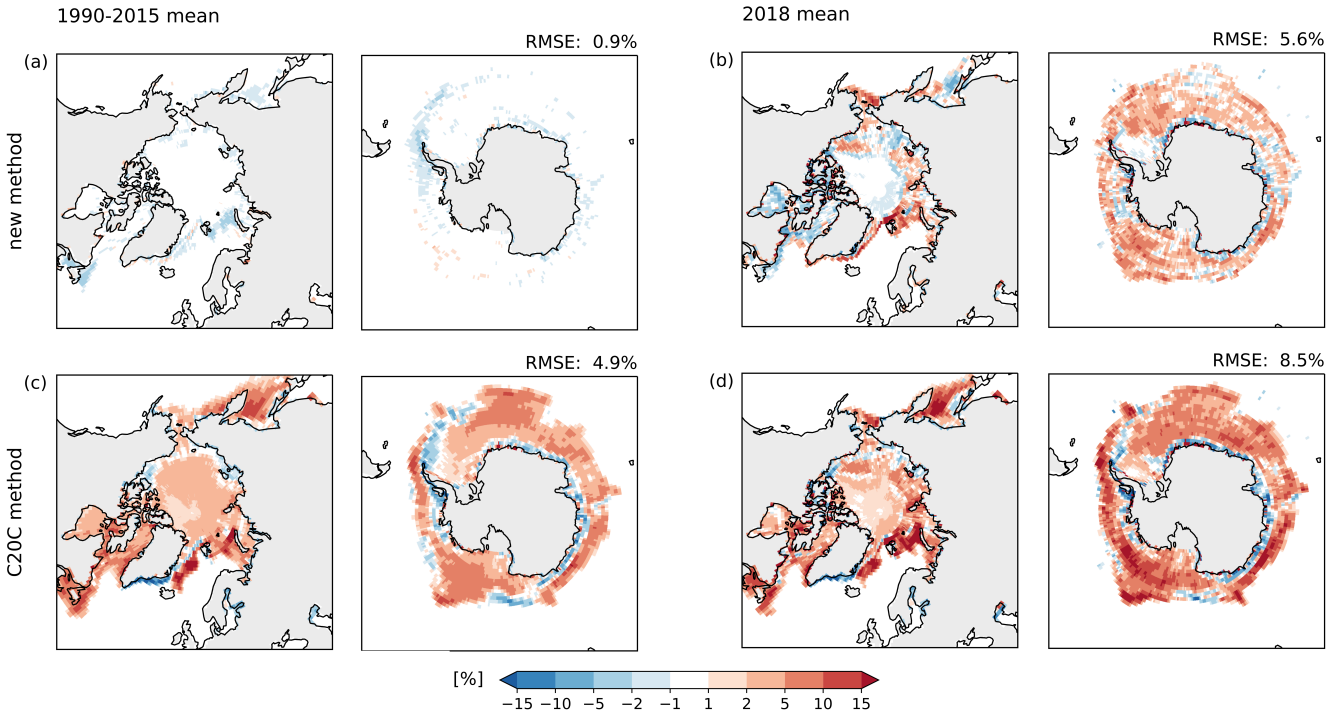

Figure A1. Evaluation of sea ice reconstruction. ~~Bias~~ Shown is the bias (%) of reconstructed historical sea ice for ~~1990–2015~~ the (~~left~~a,c) 1990–2015 mean and ~~2018~~ (~~right~~b,d) 2018 mean compared to NOAA OIv2 ~~(%)~~for areas north of 50°N and south of 50°S. The new method developed in this study (~~top~~a,b) is compared to the ~~SP equation~~C20C method (c,d). Numbers in the upper right corner indicate the global mean RMSE for grid cells that are covered by ice in the observed or estimated fields.

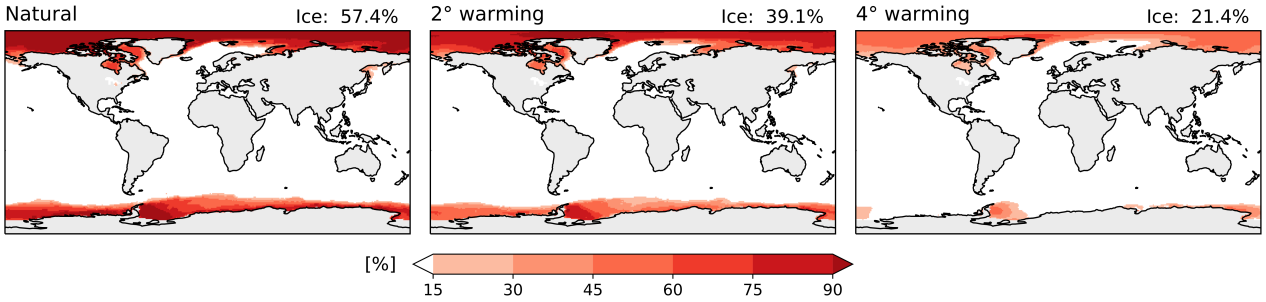

Figure A2. Absolute sea ice coverage fractions for 2018 in the natural, warming20 and warming40 simulation determined with the algorithm presented in this study. The numbers in the upper right corner indicate the yearly average sea ice coverage fraction for ice grid cells.

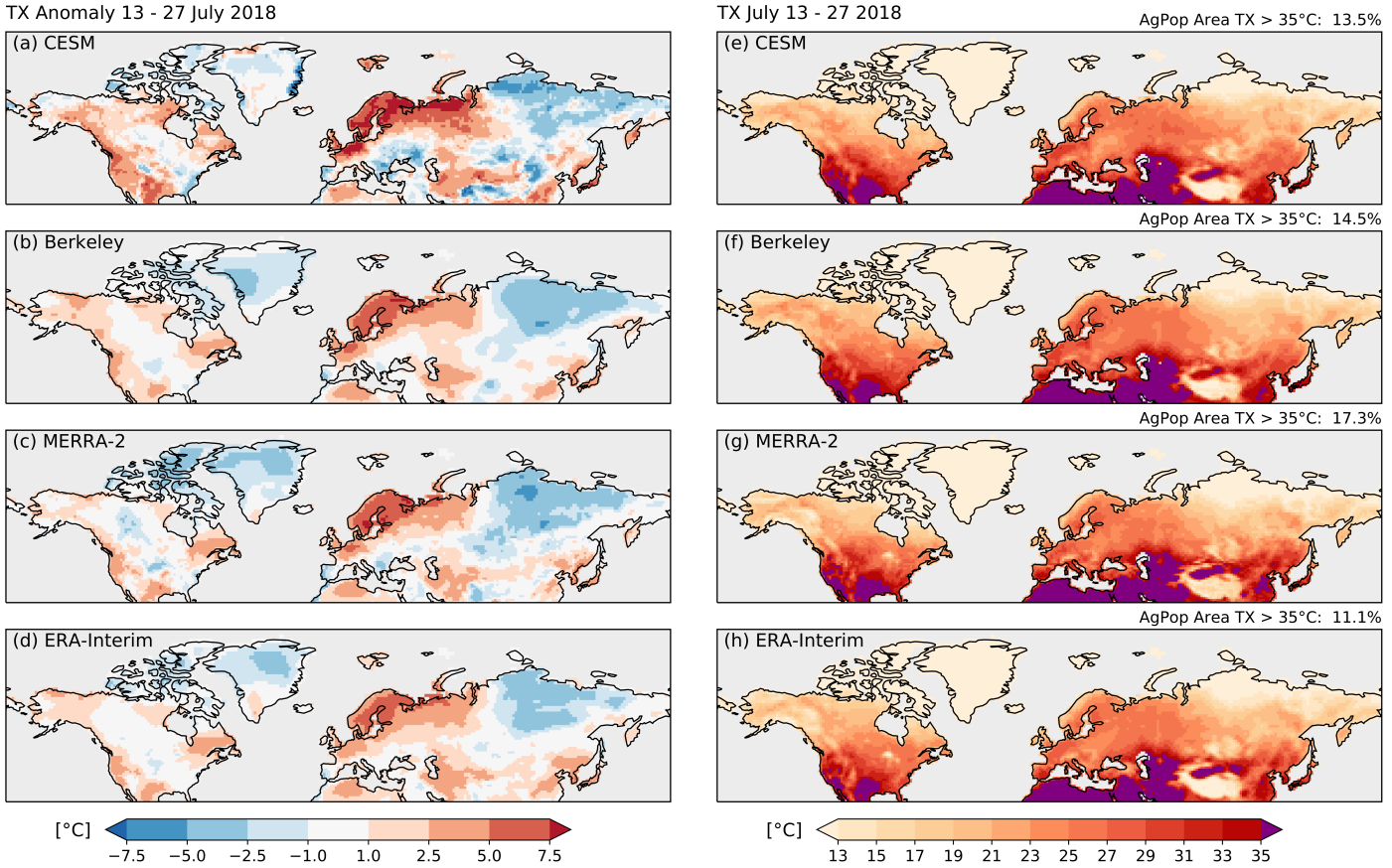

Figure A3. ~~Anomalies of maximum~~ Maximum daily temperature (TX) and anomalies of TX averaged over July ~~13 to 27 2018~~ 13–27 2018. (a) TX Anomalies are shown with respect to the 1981–2010 climatology for the historical simulations~~,~~; (b) same for ERA-Interim~~,~~; (c) MERRA-2; and (d) Berkeley. (e) Absolute TX values for CESM are bias-corrected using quantile mapping and Berkeley Earth as reference to calibrate the correction. (f) Absolute TX from Berkeley; (g) MERRA-2; and (h) ERA-Interim is shown as comparison. In the right corner above e-h the percentage of the AgPop region where TX averaged over July 13–27 2018 exceeds 35°C is given.

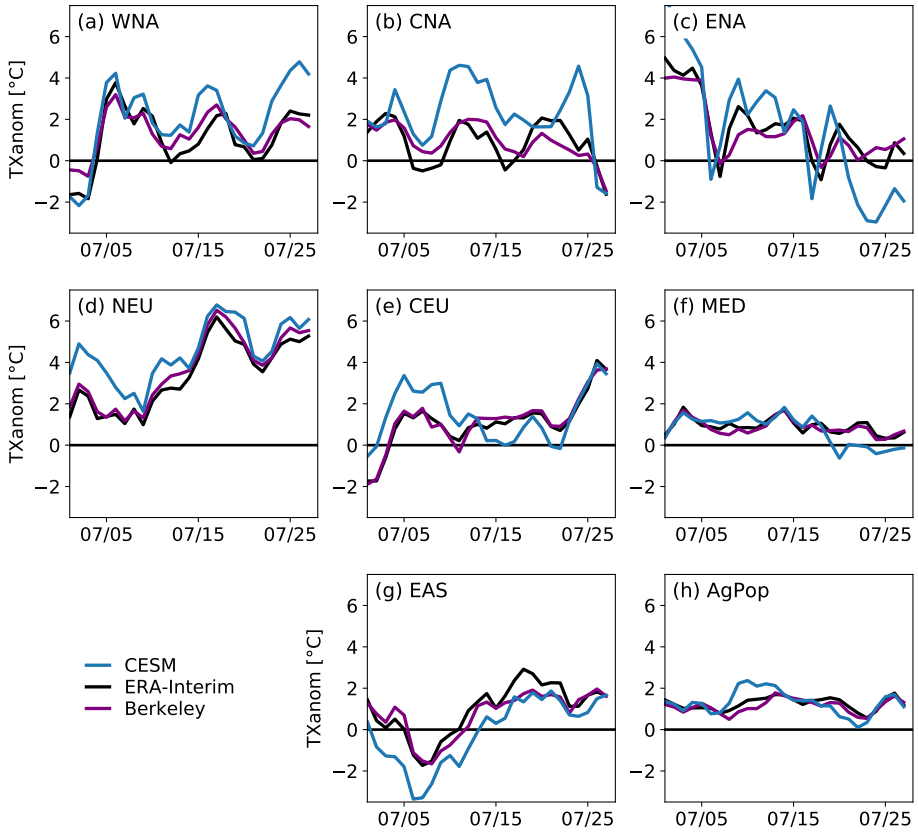

Figure A4. Time series of anomaly of maximum daily temperature (TXanom) for July 2018 averaged for (a-g) seven SREX regions and (h) the AgPop region~~and six SREX regions~~. Shown are the historical simulation from CESM, ERA-Interim and Berkeley. The reference climatology is 1981–2010.

Table A1. Slopes for all possible combinations of the observational data sets used for the linear regression as well as their uncertainties (one standard deviation).

TXreg	CRU TSv4.03		Berkeley Earth	
Tglob	GISTEMPv4	HadCRUT4	GISTEMPv4	HadCRUT4
WNA	0.85 ± 0.21	1.03 ± 0.24	0.85 ± 0.22	1.04 ± 0.25
CNA	0.16 ± 0.30	0.24 ± 0.34	0.55 ± 0.31	0.68 ± 0.35
ENA	0.56 ± 0.22	0.65 ± 0.25	0.58 ± 0.21	0.68 ± 0.23
CEU	1.52 ± 0.30	1.84 ± 0.33	1.30 ± 0.30	1.60 ± 0.34
NEU	1.13 ± 0.31	1.32 ± 0.35	1.11 ± 0.29	1.29 ± 0.33
MED	1.67 ± 0.15	1.93 ± 0.16	1.21 ± 0.16	1.40 ± 0.18
EAS	0.61 ± 0.16	0.80 ± 0.18	0.55 ± 0.19	0.75 ± 0.21
AgPop	0.72 ± 0.12	0.87 ± 0.14	0.80 ± 0.13	0.96 ± 0.14

Appendix B: Ocean forcing files

This section explains in detail how the ocean forcing files are prepared using model output for the historical and RCP8.5 scenario from the CMIP5 data archive.

B1 Step-by-step generation of delta SST and SST input files

Warming levels are determined using near-surface air temperature ("tas") from CMIP5. Delta SSTs are computed from sea surface temperature fields ("tos").

1. For easier computation later-on and to prevent steep gradients at the continent's edges, the ~~tos~~ "tos" fields are grid-filled using Poisson's equation relaxation scheme, where this has not been done by the model. All data are regridded to the 1x1 degree grid used in the NOAA OIv2 (Reynolds et al., 2002) and HadISST1 (Rayner et al., 2003) observational products, which are used to prescribe sea surface temperatures and sea ice coverage in CESM.

2. To find the years corresponding to the warming levels ~~we compute~~ weighted global yearly means of "tas" are computed for the historical and RCP8.5 time period. ~~We only include~~ Only models that provide complete monthly data for both ~~tas and tos~~ "tas" and "tos" for the historical (starting latest 1861) and the RCP8.5 time period (at least until 2099) are included.

3. ~~We define the~~ The pre-industrial reference period is defined from 1861–1880~~and average over this period~~. "Tas" for each model ~~separately. We compute the warming per model~~ is averaged over this period separately. The warming for each model is computed by taking the difference between the yearly averages from step 2 and the pre-industrial reference period. ~~We compute a~~ A 21-year running mean is taken over the yearly warming values ~~and then take~~ first, and then the multi-model mean (MMM) is computed.

4. The first year where the multi model warming ~~equals~~ exceeds 1.5°C, 2°C, 3°C and 4°C is chosen to compute the delta SST fields in the following and also to set the aerosols and GHG forcing from the RCP8.5 scenarios for the simulations. ~~We get~~The numbers in this study are: for the current warming (2018) a value of 1.12° and the 1.5°, 2°, 3° and 4° warming levels are reached in 2028, 2042, 2064, 2085 respectively, for the MMM.

5. ~~A~~ Now the delta SSTs are computed. First, a 21-year boxcar filter (running mean over months) is applied to the monthly ~~tos~~ "tos" fields for the merged historical (1975–2005) + RCP8.5 time period. These ocean fields are saved together with the multi-year monthly averaged fields for the pre-industrial time period (1861–1880).

6. The MMM is computed for the monthly ocean fields. Then the delta SST for the natural scenario is computed by subtracting the MMM pre-industrial monthly climatology from the present-day monthly fields (~~1979–2020~~2015–2018 to include spin-up). The delta SSTs for the warming scenarios are computed by subtracting the present-day monthly fields ~~(1979–2020)~~ from the MMM field of the ~~year~~ years of the warming scenario under consideration (2042 plus the years 2039–2041 for spin-up), hence:

deltaSST_natural: present $deltaSST\_natural = present - pre$ pre-industrial$industrial$; for *present* between 1979–2020 2015–2018 and *pre-industrial* averaged over 1861–1880.
deltaSST_warming20: 20warming - present; for $deltaSST\_warming20 = warming20 - present$; for *20warming*$warming20$ corresponding to year 2042 years 2039–2042 (with the 21-year boxcar filter applied to it first) and *present* between 1979–2020. 2015–2018.

Note that the delta SSTs computed this way these delta SSTs are transient.

7. We compute the SST input for the simulations simply by subtracting the natural deltaSST field and adding the warming deltaSST fields to the historical SSTs of the model. We The constraint of Hurrell et al. (2008) is used to ensure that temperature is not below $-1.8°C$.

Author contributions. KW, MH and SIS designed the experiments and discussed the results. KW developed the sea ice reconstruction method with input by MH and based on code used for HAPPI that was made available by Eunice Lo from University of Bristol. KW ran the model simulation and analysed the results. KW prepared the manuscript with contributions from all co-authors

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank Eunice Lo for sharing the HAPPI sea ice code. We acknowledge also thank Urs Beyerle and Richard Wartenburger for retrieving the CMIP5 data. We are grateful to Geert Jan van Oldenborgh and an anonymous reviewer for evaluating our manuscript and for their helpful comments. We acknowledge financial support from the European Research Council (ERC) "DROUGHT-HEAT" project (Grant 617518) funded by the European Community's Seventh Framework Programme. KW thanks Sven Kotlarski for his input on bias correction and quantile mapping.

References

Alter, R. E., Douglas, H. C., Winter, J. M., and Eltahir, E. A. B.: Twentieth Century regional climate change during the summer in the Central United States attributed to agricultural intensification, Geophysical Research Letters, 45, 1586–1594, https://doi.org/10.1002/2017GL075604, 2018.

Barriopedro, D., Sousa, P. M., Trigo, R. M., García-Herrera, R., and Ramos, A. M.: The Exceptional Iberian Heatwave of Summer 2018, Bulletin of the American Meteorological Society, 101, S29–S34, https://doi.org/10.1175/BAMS-D-19-0159.1, 2020.

Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Quarterly Journal of the Royal Meteorological Society, 140, 1935–1944, https://doi.org/10.1002/qj.2297, 2014.

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M. E., Miller, S. K., Steinman, B. A., Stolpe, M. B., and Way, R. G.: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, Geophysical Research Letters, 42, 6526–6534, https://doi.org/10.1002/2015GL064888, 2015.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values, Global and Planetary Change, 57, 16–26, https://doi.org/10.1016/j.gloplacha.2006.11.030, 2007.

Dlugokencky, E.: NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends_ch4/, accessed 16 October 2019, 2019.

Dlugokencky, E. and Tans, P.: NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends/, accessed 16 October 2019, 2019.

Donat, M. G., Pitman, A. J., and Seneviratne, S. I.: Regional warming of hot extremes accelerated by surface energy fluxes, Geophysical Research Letters, 44, 7011–7019, https://doi.org/10.1002/2017GL073733, 2017.

Drouard, M., Kornhuber, K., and Woollings, T.: Disentangling dynamic contributions to summer 2018 anomalous weather over Europe, Geophysical Research Letters, 46, 12 537–12 546, https://doi.org/10.1029/2019GL084601, 2019.

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., and Schär, C.: Soil moisture–atmosphere interactions during the 2003 European Summer Heat Wave, Journal of Climate, 20, 5081–5099, https://doi.org/10.1175/JCLI4288.1, 2007.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

GISTEMP Team: GISS Surface Temperature Analysis (GISTEMP), version 4, NASA Goddard Institute for Space Studies. https://data.giss.nasa.gov/gistemp/, accessed 4 July 2020, 2020.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Hazeleger, W., van den Hurk, B. J. J. M., Min, E., van Oldenborgh, G. J., Petersen, A. C., Stainforth, D. A., Vasileiadou, E., and Smith, L. A.: Tales of future weather, Nature Climate Change, 5, 107–113, https://doi.org/10.1038/nclimate2450, 2015.

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., and Rosinski, J.: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model, Journal of Climate, 21, 5145–5153, https://doi.org/10.1175/2008JCLI2292.1, 2008.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Jones, P. W.: First- and second-order conservative remapping schemes for grids in spherical coordinates, Monthly Weather Review, 127, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

KMA: 2018 Annual Report, Korea Meteorological Administration, (available through: https://www.kma.go.kr/download_01/Annual_Report_2018.pdf), 2019.

KNMI: Zomer 2018 (juni, juli, augustus): extreem warm, zeer zonnig en zeer droog, Koninklijk Nederlands Meteorologisch Instituut, (available through: https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2018/zomer, 2020.

Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., and Russell, L. M.: Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD018588, d23204, 2012.

Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J. F., Lehmann, J., and Horton, R. M.: Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions, Nature Climate Change, https://doi.org/10.1038/s41558-019-0637-z, 2019a.

Kornhuber, K., Osprey, S., Coumou, D., Petri, S., Petoukhov, V., Rahmstorf, S., and Gray, L.: Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern, Environmental Research Letters, 14, https://doi.org/10.1088/1748-9326/ab13bf, 2019b.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, Journal of Advances in Modeling Earth Systems, 3, https://doi.org/10.1029/2011MS00045, 2011.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, Journal of Geophysical Research: Atmospheres, 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., Clilverd, M. A., Dudok de Wit, T., Haberreiter, M., Hendry, A., Jackman, C. H., Kretzschmar, M., Kruschke, T., Kunze, M., Langematz, U., Marsh, D. R., Maycock, A. C., Misios, S., Rodger, C. J., Scaife, A. A., Seppälä, A., Shangguan, M., Sinnhuber, M., Tourpali, K., Usoskin, I., van de Kamp, M., Verronen, P. T., and Versick, S.: Solar forcing for CMIP6 (v3.2), Geoscientific Model Development, 10, 2247–2302, https://doi.org/10.5194/gmd-10-2247-2017, 2017.

Mitchell, D., AchutaRao, K., Allen, M., Bethke, I., Beyerle, U., Ciavarella, A., Forster, P. M., Fuglestvedt, J., Gillett, N., Haustein, K., Ingram, W., Iversen, T., Kharin, V., Klingaman, N., Massey, N., Fischer, E., Schleussner, C.-F., Scinocca, J., Seland, Ø., Shiogama, H., Shuckburgh, E., Sparrow, S., Stone, D., Uhe, P., Wallom, D., Wehner, M., and Zaaboul, R.: Half a degree additional warming, prognosis and projected impacts (HAPPI): Background and experimental design, Geoscientific Model Development, 10, 571–583, https://doi.org/10.5194/gmd-10-571-2017, 2017.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M., Smith, A. K., Tilmes, S., Vitt, F., Morrison, H., Cameron-Smith, P., Collins, W. D., Iacono, M. J., Easter, R. C., Ghan, S. J., Liu, X., Rasch, P. J., and Taylor, M. A.: Description of the NCAR Community Atmosphere Model (CAM 5.0), Technical report, National Center for Atmospheric Research. Boulder, Colorado, 2012.

NOAA: State of the climate: Global climate report for July 2018, National Centers for Environmental Information, https://www.ncdc.noaa.gov/sotc/global/201807, 2018.

NOAA Earth System Research Laboratory: Combined nitrous oxide data from the NOAA/ESRL Global Monitoring Division, ftp://ftp.cmdl.noaa.gov/hats/n2o/combined/HATS_global_N2O.txt, accessed 16 October 2019, 2019.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Thornton, P. E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Hoffman, F., Lamarque, J.-F., Mahowald, N., Niu, G.-Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stöckli, R., Wang, A., Yang, Z.-L., Zeng, X., and Zeng, X.: Technical description of version 4.0 of the Community Land Model (CLM), Technical report, National Center for Atmospheric Research. Boulder, Colorado, 2010.

Rajczak, J., Kotlarski, S., Salzmann, N., and Schär, C.: Robust climate scenarios for sites with sparse observations: a two-step bias correction approach, International Journal of Climatology, 36, 1226–1243, https://doi.org/10.1002/joc.4417, 2016.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002670, 2003.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, Journal of Climate, 15, 1609–1625, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2, 2002.

Richardson, M., Cowtan, K., Hawkins, E., and Stolpe, M. B.: Reconciled climate response estimates from climate models and the energy budget of Earth, Nature Climate Change, 6, 931–935, https://doi.org/10.1038/nclimate3066, 2016.

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C.: A new estimate of average Earth surface land temperature spanning 1743 to 2011, Geoinformatics & Geostatistics: An Overview, 1, https://doi.org/10.4172/2327-4581.1000101, 2013a.

Rohde, R., Muller, R. A., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., and Mosher, S.: Berkeley Earth temperature averaging process, Geoinformatics & Geostatistics: An Overview, 1, https://doi.org/10.4172/2327-4581.1000103, 2013b.

Seneviratne, S. I. and Hauser, M.: Regional climate sensitivity of climate extremes in CMIP6 vs CMIP5 multi-model ensembles, Earth's Future, https://doi.org/10.1029/2019EF001474, 2020.

Seneviratne, S. I., Luthi, D., Litschi, M., and Schar, C.: Land-atmosphere coupling and climate change in Europe, Nature, 443, 205–209, https://doi.org/10.1038/nature05095, 2006.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., Innes, K. M., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., Zhang, X., Rusticucci, M., Semenov, V., Alexander, L. V., Allen, S., Benito, G., Cavazos, T.,

Clague, J., Conway, D., Della-Marta, P. M., Gerber, M., Gong, S., Goswami, B. N., Hemer, M., Huggel, C., den Hurk, B. V., Kharin, V. V., Kitoh, A., Tank, A. M. G. K., Li, G., Mason, S., Guire, W. M., Oldenborgh, G. J. V., Orlowsky, B., Smith, S., Thiaw, W., Velegrakis, A., Yiou, P., Zhang, T., Zhou, T., and Zwiers, F. W.: Changes in climate extremes and their impacts on the natural physical environment, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, pp. 109–230, edited by C. B. Field et al., Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781139177245.006, 2012.

Seneviratne, S. I., Phipps, S. J., Pitman, A. J., Hirsch, A. L., Davin, E. L., Donat, M. G., Hirschi, M., Lenton, A., Wilhelm, M., and Kravitz, B.: Land radiative management as contributor to regional-scale climate adaptation and mitigation, Nature Geoscience, 11, 88–96, https://doi.org/10.1038/s41561-017-0057-5, 2018.

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., Fowler, H. J., James, R., Maraun, D., Martius, O., Senior, C. A., Sobel, A. H., Stainforth, D. A., Tett, S. F. B., Trenberth, K. E., van den Hurk, B. J. J. M., Watkins, N. W., Wilby, R. L., and Zenghelis, D. A.: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change, Climatic Change, 151, 555–571, https://doi.org/10.1007/s10584-018-2317-9, 2018.

Sippel, S., Zscheischler, J., Mahecha, M. D., Orth, R., Reichstein, M., Vogel, M., and Seneviratne, S. I.: Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics, Earth System Dynamics, 8, 387–403, https://doi.org/10.5194/esd-8-387-2017, 2017.

Stone, D. A., Christidis, N., Folland, C., Perkins-Kirkpatrick, S., Perlwitz, J., Shiogama, H., Wehner, M. F., Wolski, P., Cholia, S., Krishnan, H., Murray, D., Angélil, O., Beyerle, U., Ciavarella, A., Dittus, A., Quan, X.-W., and Tadross, M.: Experiment design of the International CLIVAR C20C+ Detection and Attribution project, Weather and Climate Extremes, 24, https://doi.org/10.1016/j.wace.2019.100206, 2019.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

Thiery, W., Visser, A. J., Fischer, E. M., Hauser, M., Hirsch, A. L., Lawrence, D. M., Lejeune, Q., Davin, E. L., and Seneviratne, S. I.: Warming of hot extremes alleviated by expanding irrigation, Nature Communications, 11, https://doi.org/10.1038/s41467-019-14075-4, 2020.

Toreti, A., Belward, A., Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie, O., Seguini, L., Manfron, G., Lopez-Lozano, R., Baruth, B., van den Berg, M., Dentener, F., Ceglar, A., Chatzopoulos, T., and Zampieri, M.: The exceptional 2018 European water seesaw calls for action on adaptation, Earth's Future, 7, 652–663, https://doi.org/10.1029/2019EF001170, 2019.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: An overview, Climatic Change, 109, https://doi.org/10.1007/s10584-011-0148-z, 2011.

Vogel, M. M., Zscheischler, J., and Seneviratne, S. I.: Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe, Earth System Dynamics, 9, 1107–1125, https://doi.org/10.5194/esd-9-1107-2018, 2018.

Vogel, M. M., Zscheischler, J., Wartenburger, R., Dee, D., and Seneviratne, S. I.: Concurrent 2018 hot extremes across Northern Hemisphere due to human-induced climate change, Earth's Future, 7, 692–703, https://doi.org/10.1029/2019EF001189, 2019.

Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., and Seneviratne, S. I.: Assessing the dynamic versus thermodynamic origin of climate model biases, Geophysical Research Letters, 45, 8471–8479, https://doi.org/10.1029/2018GL079220, 2018.

595 Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., and Seneviratne, S. I.: Identifying key driving processes of major recent heat waves, Journal of Geophysical Research: Atmospheres, 124, 11 746–11 765, https://doi.org/10.1029/2019JD030635, 2019.

Wuebbles, D. J., Fahey, D. W., Hibbard, K. A., Dokken, D. J., Stewart, B. C., Maycock, T. K., Easterling, D. R., Kunkel, K. E., Arnold, J. R., Knutson, T., LeGrande, A. N., Leung, L. R., Vose, R. S., Waliser, D. E., and Wehner, M. F.: Precipitation change in the United States, pp. 207–230, U.S. Global Change Research Program, Washington, DC, USA, https://doi.org/10.7930/J0H993CC, 2017.

600 Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P.-L., Rasch, P. J., Neubauer, D., and Lohmann, U.: Technical note: On the use of nudging for aerosol–climate model intercomparison studies, Atmospheric Chemistry and Physics, 14, 8631–8645, https://doi.org/10.5194/acp-14-8631-2014, 2014.