Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL082062

Key Points:
- We detect a clear climate change signal in extreme heat, heat stress, and cold over Europe that cannot be explained by internal variability
- On average across Europe, days with extreme heat and heat stress have tripled and days with extreme cold more than halved from 1950–2018
- Hot and cold extremes warmed significantly more than the corresponding seasonal mean in Central Europe, by 2.3 and >3 °C, respectively

Supporting Information:
- Supporting Information S1

Correspondence to:
R. Lorenz,
ruth.lorenz@env.ethz.ch

Citation:
Lorenz, R., Stalhandske, Z., & Fischer, E. M. (2019). Detection of a climate change signal in extreme heat, heat stress, and cold in Europe from observations. *Geophysical Research Letters*, *46*, 8363–8374. https://doi.org/10.1029/2019GL082062

Received 15 JAN 2019
Accepted 8 JUL 2019
Accepted article online 17 JUL 2019
Published online 29 JUL 2019

Detection of a Climate Change Signal in Extreme Heat, Heat Stress, and Cold in Europe From Observations

Ruth Lorenz[1], Zélie Stalhandske[1], and Erich M. Fischer[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

Abstract In the last two decades Europe experienced a series of high-impact heat extremes. We here assess observed trends in temperature extremes at ECA&D stations in Europe. We demonstrate that on average across Europe the number of days with extreme heat and heat stress has more than tripled and hot extremes have warmed by 2.3 °C from 1950–2018. Over Central Europe, the warming exceeds the corresponding summer mean warming by 50%. Days with extreme cold temperatures have decreased by a factor of 2–3 and warmed by more than 3 °C, regionally substantially more than winter mean temperatures. Cold and hot extremes have warmed at about 94% of stations, a climate change signal that cannot be explained by internal variability. The clearest climate change signal can be detected in maximum heat stress. EURO-CORDEX RCMs broadly capture observed trends but the majority underestimates the warming of hot extremes and overestimates the warming of cold extremes.

1. Introduction

Human activities are estimated to have caused approximately 1.0 °C of global warming above preindustrial levels (Masson-Delmotte et al., 2018). On top of the mean temperature warming, a trend to more frequent and intense heat extremes has been observed (Alexander et al., 2006; Coumou & Robinson, 2013; Della-Marta et al., 2007). During the last decades Europe experienced a series of heat extremes that broke long-standing temperature records almost everywhere (Barriopedro et al., 2011; King, 2017) with summer 2018 being very warm over most of Europe. Europe turns out to be one of the areas that experienced the strongest intensification of hot extremes since the 1950s (Donat et al., 2013; Fischer & Knutti, 2014). At the same time there has been a substantial decrease in the frequency of cold extremes (Alexander, 2016; Frich et al., 2002).

From a statistical perspective temperature extremes can change as a result of mean warming or changes in variance and skewness of the distribution (Katz & Brown, 1992; Loikith et al., 2018; Mearns et al., 1984; Schär et al., 2004; Seneviratne et al., 2012). Model projections suggest that temperature variability will regionally decrease in winter and increase in summer (Cattiaux et al., 2015; Fischer et al., 2011; Holmes et al., 2016; Kjellström et al., 2007; Orlowsky & Seneviratne, 2012), leading to cold and warm extremes warming stronger than the mean. However, the observational evidence for variability changes is limited, and significant changes in summer temperature variability have been documented only at few stations in central and southern Europe (Della-Marta et al., 2007; Yiou et al., 2009). Recently, Gross et al. (2019) found based on a gridded temperature data set the largest differences between extreme and seasonal mean warming rates in the cold tails of the distributions for many regions in the extratropical Northern Hemisphere and the smallest differences for boreal summer. Thus, differences in changes between extremes and seasonal means are heterogeneous in space and time. We here test for a large network of stations whether changes observed in extreme temperature are only due to a shift in mean temperatures toward a globally warmer climate or if there is also a change in the variability of the temperature distribution.

Trend detection of extremes has mostly been done at global to hemispheric scale. At regional to local scale internal variability can strongly offset or amplify local to regional trends in extremes over several decades (Fischer & Knutti, 2013; Perkins & Fischer, 2013). We here test whether we can detect a climate change signal at stations across Europe. In order to minimize the effect of internal variability, it is essential to aggregate across large regions and analyze as long periods as possible. Likewise, we also reduce the sensitivity to potential inhomogeneities at individual stations. Finally, most observational studies focus on temperature only, whereas we here account for the effect of ambient humidity on heat stress, which is potentially rele-

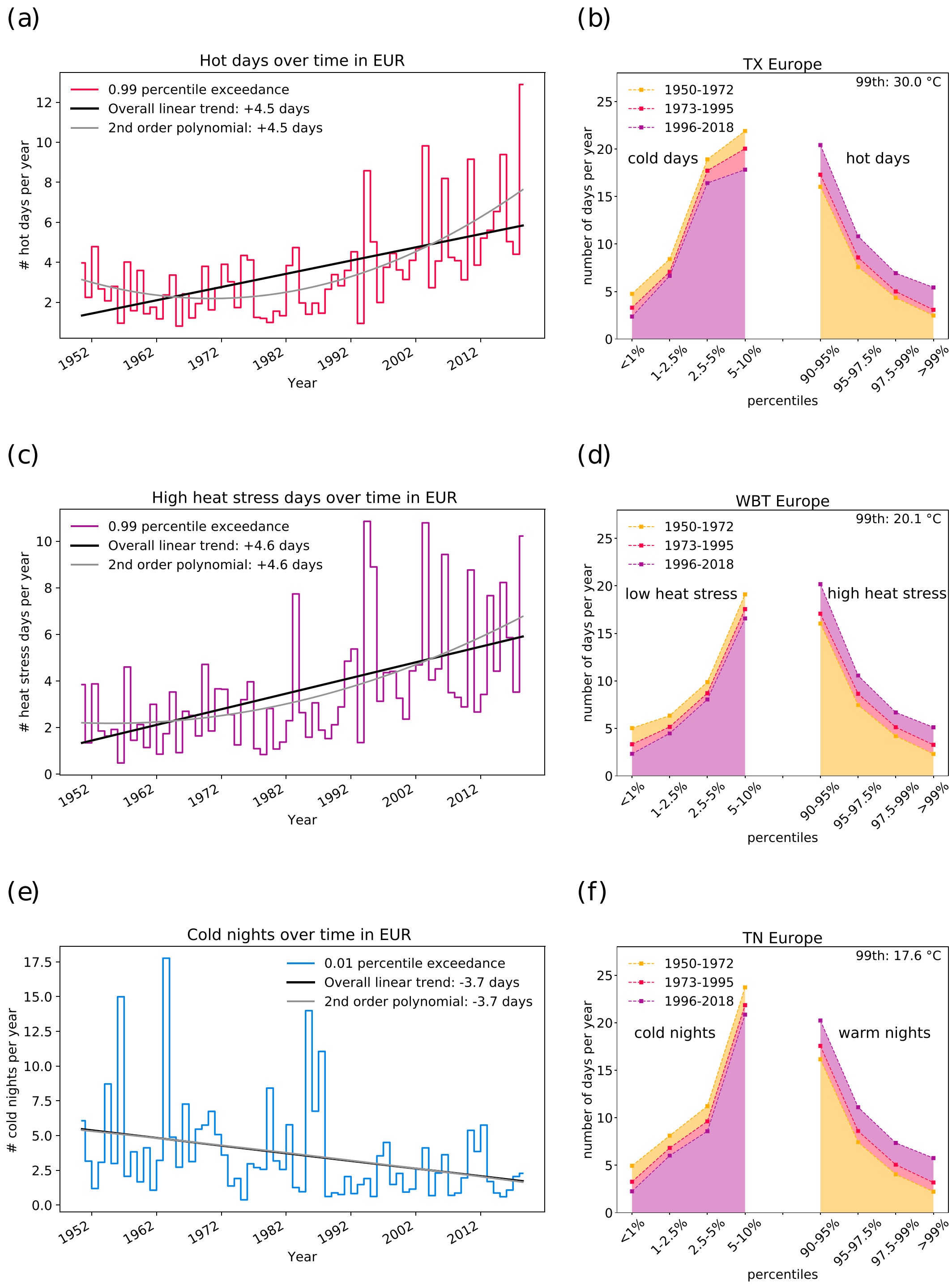

Figure 1. Frequency of hot (a, b), heat stress (c, d), and cold (e, f) extremes over time. The left column shows the number of days (a) TX > 99th percentile, (c) WBT > 99th percentile, and (e) TN < 1st percentile from 1950–2018 for region EUR. For illustration we added a linear trend and a second-order polynomial. The right column shows the number of days which fall within certain percentile bins for three different time periods, 1950–1972 (orange), 1973–1995 (red), and 1996–2018 (purple). The percentiles were calculated using the whole time period 1950–2018. Be aware that the bins are unequally spaced and the middle of the distribution was excluded in the figure.

vant for health and labor productivity (Diffenbaugh et al., 2007; Fischer & Schär, 2010; Pal & Eltahir, 2016). Most previous studies quantifying changes in heat stress were based on climate models. We here address the question whether a signal in heat stress can be detected also in Europe.

Thus, we aim here at revisiting and detecting trends in temperature extremes for Europe. We assess how the recent years fit the earlier trends and the projections. We first quantify observed trends in the frequency and intensity of temperature extremes and test whether a change can be clearly detected, that is, whether they are larger than expected from internal variability. We also include wet-bulb temperature, which combines humidity and air temperature, as a measure for heat stress. In addition, we test if changes in the extremes are larger than changes in the mean, as has been proposed in multiple modeling studies. Finally, we assess how well EURO-CORDEX regional climate models (RCMs) reproduce the observed trends.

2. Data and Methods

2.1. Data Sets

We use daily mean (TG), daily maximum (TX), daily minimum temperature (TN), and daily relative humidity (HU) from the European Climate Assessment & Dataset (ECA&D, Klein Tank et al., 2002). This data set provides quality-controlled station data for around 4,000 stations in Europe, the exact number depends on the variable. Data are available from 1950 to present, we use data until October 2018.

We consider a station to be valid if at least 90% of the data is available over the whole time period 1950–2018 and for Figure 1 if at least 80% of data is available for each of the three subperiods. Every year needs at least 300 days valid data, otherwise we excluded this year when calculating yearly maxima or minima. When calculating seasonal means we allowed not more than ≈10% of data to be missing per season (>82 days). These constraints reduce the data set to around 1,000 stations for temperatures and even less for humidity (~440). However, these constraints are necessary to avoid spurious trends due to missing data in the time series. For instance, imagine a station which only includes data until January 2017. Since we are looking for maximum values per year, including this year would lead to very low yearly maxima in 2017. The wet-bulb temperature (WBT) we calculate from humidity and mean temperature using the empirical equation by Stull, (2011; more details in Text S1 in the supporting information). We group valid stations into four different groups shown in Figure 2 (left panels), Northern Europe (NEU), Central Europe (CEU), Mediterranean (MED) (Seneviratne et al., 2012), and Europe (EUR).

We also use the gridded version E-Obs v19e (Cornes et al., 2018) of the same temperature data and RCM output from EURO-CORDEX (0.44° resolution; Kotlarski et al., 2014). The E-Obs gridded data set covers the period 1950–2018. Some of the EURO-CORDEX runs are only available from 1971 onward. Hence, we only use the time period 1971–2018 for the E-Obs versus EURO-CORDEX comparison.

2.2. Statistical Methods

We investigate yearly maxima and minima and calculate 1-day, 3-day, 5-day, and 7-day means (rolling means, centered day of interest) for summer (June, July, August) and winter (December, January, February). For TX and WBT we extract the yearly maximum, denoted TXx and WBTx (1-day), TX3x and WBT3x (3-day), and so forth. For TN we look for the yearly minimum (TNn, TN3n, TN5n, and TN7n). In addition, we calculate seasonal means of TG for each year for summer (TG_{JJA}) and winter (TG_{DJF}). Anomalies were calculated by subtracting the seasonal mean over the whole 1950–2018 time period. Then we estimate the linear trends based on anomalies using least-squares linear regressions.

We perform block bootstrapping (Wilks, 1997) over all stations at the same time (details in Text S2 in the supporting information) to estimate trends expected due to internal variability. To assess if trends at individual stations are significant we use a two-sided *p* value (Wald-test with *t* distribution) adjusted using the false discovery rate (FDR; Wilks, 2006, 2016) to account for multiple testing. We calculate the differences in trend between $TXx-TG_{JJA}$ and $TNn-TG_{DJF}$ at individual stations and compare this distribution to a bootstrapped distribution (Figure 3). To assess if the median trend change in the extremes is statistically different than the seasonal mean trend we use a two-sided *t* test.

3. Results

3.1. More Hot and Less Cold Extremes

Days with extreme heat (TX > 99th percentile) as well as extreme heat stress (WBT > 99th percentile) have at least tripled over the period 1950–2018. On average across Europe (EUR region) they increased from around

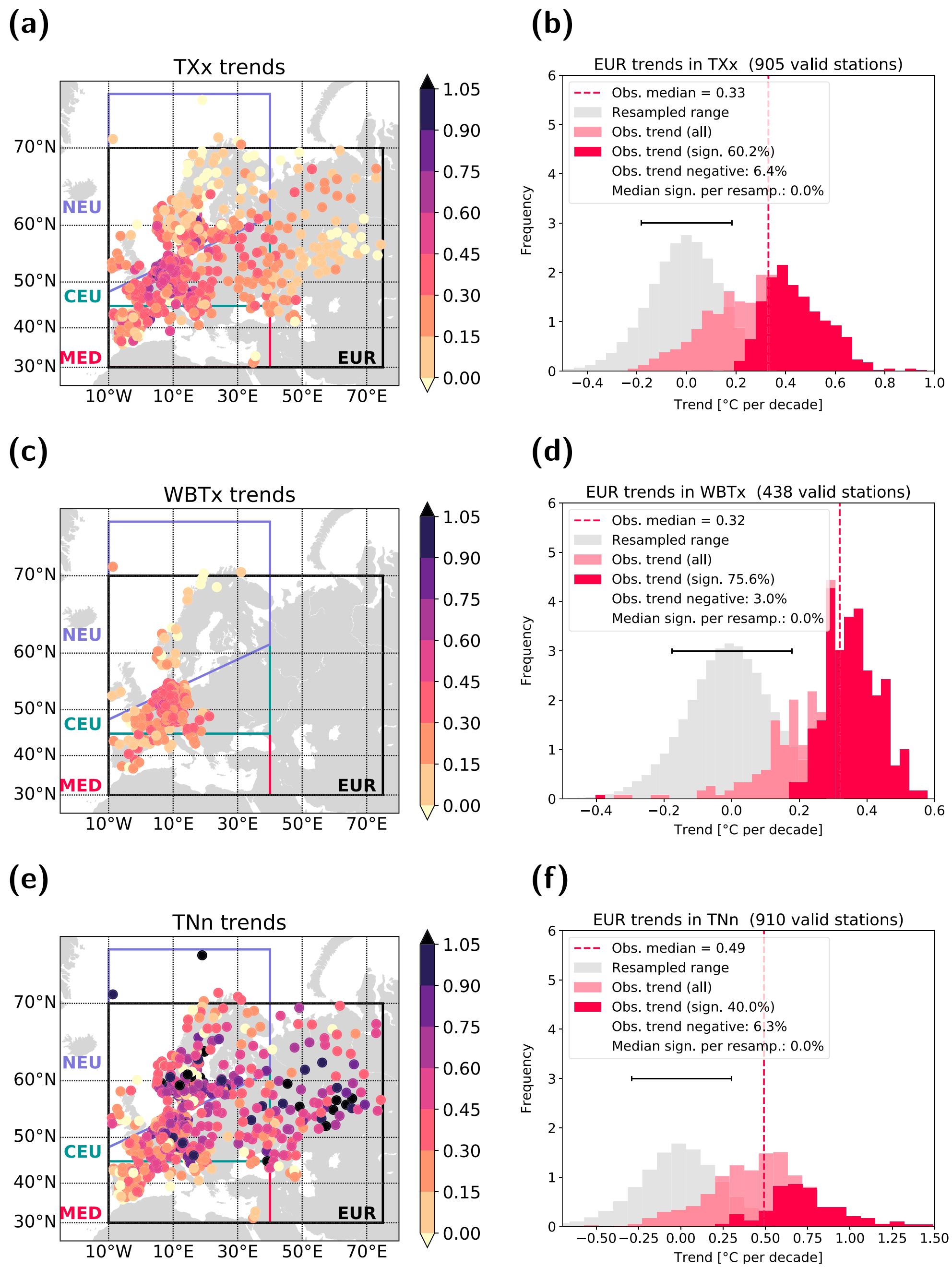

Figure 2. The left column shows trends in (a) TXx, (c) WBTx, and (e) TNn at valid stations. In addition, the other subregions used in the analysis are indicated on the map. The right column shows histograms of trends in (b) TXx, (d) WBTx, and (f) TNn in light red (all stations) and dark red (only stations with statistically significant trends) and from randomly bootstrapped time series in gray for EUR. The black range indicates the 5th to 95th percentile of the median area averaged trends from the bootstrapped samples.

2 days/year in 1950 to about 6 days/year in 2018 as estimated from a linear trend (Figures 1a and 1c). The change in the number of days exceeding a certain percentile should not necessarily be expected to be linear; therefore, we also added the change estimated from a second-order polynomial trend. Even though these two trend lines are different, the calculated increase over the time series is the same. The right side of Figure 1 shows the number of days exceeding a certain percentile for three different time periods (note that no trends were calculated for these shorter subperiods). While the increase in hot days is small from 1950–1972 to 1973–1995, days with extreme heat have doubled up to 1996–2018 and more than tripled over the entire

10.1029/2019GL082062

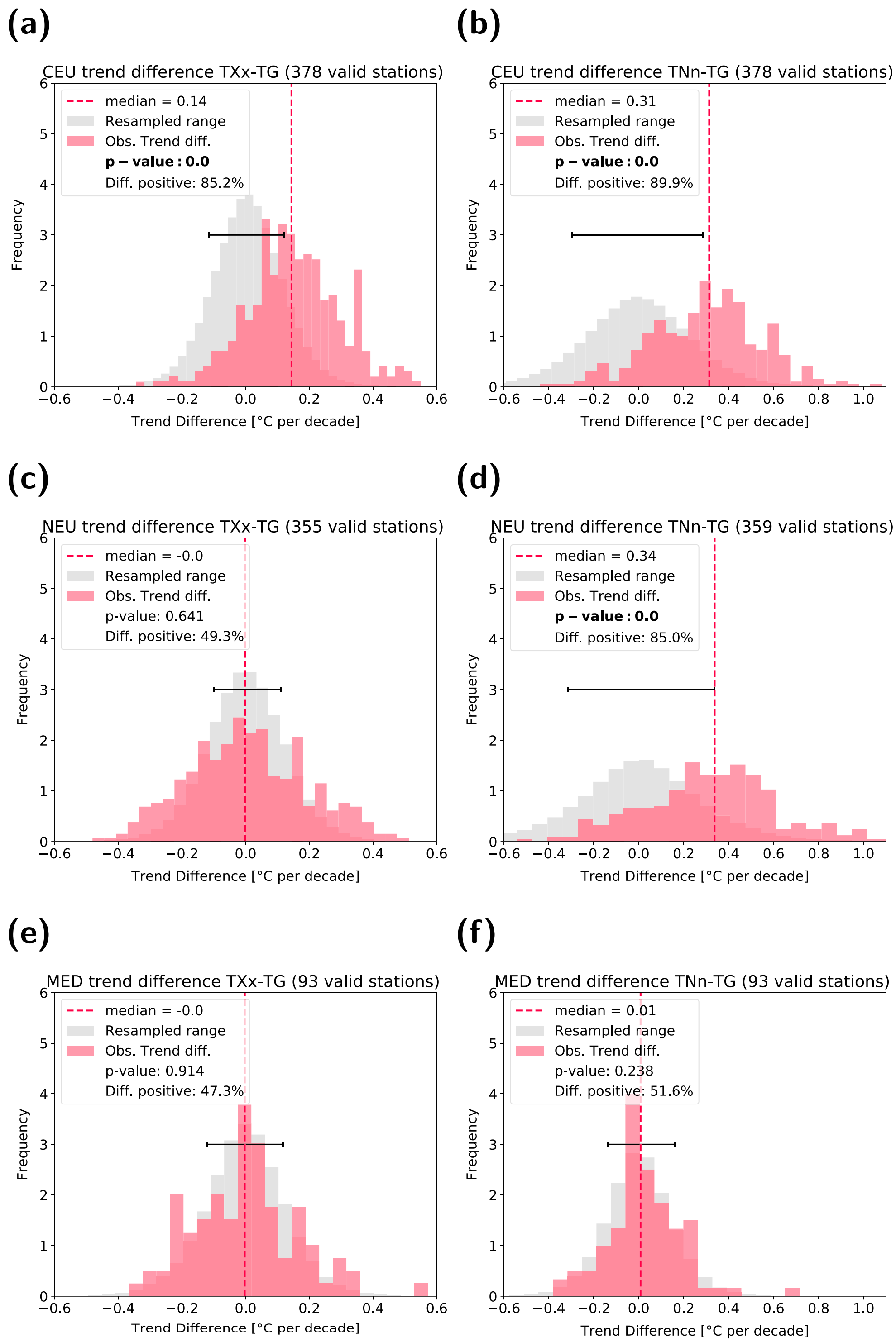

Figure 3. Histograms of the difference between TXx and TG_{JJA} (left column), and TNn and TG_{DJF} (right column) in red and differences in trends from randomly bootstrapped time series in gray for different regions (a, b) CEU, (c, d) NEU, (e, f) MED. The black range indicates the 5th to 95th percentiles of the median-averaged trends from the bootstrapped samples. The *p* value compares station data and bootstrapped distributions using a *t* test.

period (Figures 1a and 1b). The magnitude of this increase depends on the extremeness of the definition, and is largest for the most extreme bins >99% (Figures 1b and 1d). The sign of the change is robust across different percentile thresholds. Changes are consistent showing more frequent extreme hot days and warm nights across all bins of temperature. The increase in extremely warm nights is even more pronounced than for days, with more than doubling from the earliest to the most recent period (Figure 1f). Likewise, days with extreme heat stress increased consistently with day and nights of extreme heat (Figure 1d).

On the other hand, extremely cold nights (TN < 1st percentile) have decreased by a factor of 2 to 3 from more than 5 around 1950 to around 2 days/year in 2018 (Figure 1e). Again, the relative decrease is strongest for the most extreme events. During the last three decades there was no single winter with an anomalously high number of cold nights averaged across Europe. Both frequency of hot and cold extremes show considerable year-to-year variations and multidecadal trends show spatial heterogeneities partly due to internal variability. However, if aggregated across the whole of Europe, a clear signal emerges with strong trend toward more days and nights with extreme heat and heat stress and less days and nights with extreme cold.

3.2. Amplified Warming of Hot and Cold Extremes

Figure 2 shows observed trends in the intensity of hot (TXx), heat stress (WBTx) and cold (TNn) extremes at station level as well as aggregated over all stations in EUR as histograms. Hot extremes have warmed at 94% of all stations with significant trends at 60% and a median warming of 0.33 °C per decade or 2.3 °C over the period 1950–2018 across all stations in Europe (Figure 2b). In CEU the warming was up to 0.8 °C per decade at individual stations (Figure 2a) that is more than 5 °C across the whole period. Heat stress extremes (WBTx) have significantly intensified at 75% of all stations with a median trend of 0.32 °C per decade (Figures 2c and 2d). Note that the WBTx trends are limited to a much smaller network of stations mostly in western Europe that provided the necessary humidity measurements. In addition, we calculated trends for the hottest multiday extreme episodes (hottest consecutive 3, 5, and 7 days), a metric for heatwave intensity. We find that the hottest week and the week with the highest heat stress have intensified at about the same rate as the 1-day extremes (Figures S1 and S2 in the supporting information). This is relevant particularly for those impacts that only manifest themselves after a period of sustained heat (stress). Scherrer et al. (2016) even found longer hot extremes (TX7x) to warm faster than shorter hot extremes (TXx) in Switzerland. While we see a tendency for longer extremes to warm faster than shorter extremes in some regions this is not the case in others and these differences are very small (see Figures S1–S3). The difference may either relate to the different regional coverage or the longer time period (1900–2015) used in Scherrer et al. (2016). Overall, these multiday trends are very similar and consistent to the one day hottest days per year. Therefore, we concentrate on one-day extremes TXx and WBTx in the following.

Even though trends are negative at few individual stations, a very clear pattern emerges across Europe. The trend distribution across all stations is clearly different than expected by chance. A first indication is that 94% of the stations show positive and only 6% negative trends, whereas in the long run the two should balance in the absence of change and if the stations were independent. Since this is not the case we further test significance by using block bootstrapping and thereby accounting for temporal and spatial autocorrelation. The gray distributions in Figures 2b, 2d, and 2f show trends obtained by block bootstrapping as measure of what would be possible by chance. The median of the observed trend distribution is clearly outside the confidence range of medians across the block bootstrapped distributions. We also test the significance of the trends at each individual station and adjust these *p* values for multiple testing using the false discovery rate (see section 2.2; Wilks, 2006). For TXx 60% of stations show statistically significant trends, for WBTx 75% of stations. By chance and assuming spatial independence, we would only expect 2.5% of the stations to show significant positive trends (given that we are testing on a 5% level using a two-sided test). Therefore, a clear intensification of extreme heat and extreme heat stress can be detected across Europe with a distribution of trends that is very different to what would be expected from internal variability. The overall trends in heat stress are even more significantly different from internal variability, which is consistent with arguments that the signal-to-noise ratio in models and the detectability in observations is higher for heat stress metrics combining temperature and humidity (Knutson & Ploshay, 2016).

To investigate if warming in the hot extremes has been larger than in the mean, we look at the difference between the trends in hot extremes (TXx) and summer mean temperatures and plot these trend differences as histograms over all stations within a region (Figure 3, red). We test whether this warming difference between hot extremes and summer mean is different to what is expected by chance. To this end we block bootstrap TXx and TG_{JJA} and plot the distributions of their difference in gray, where this difference in the mean is zero.

We find that in CEU heat extremes warmed more than the summer mean at about 85% of the stations and with a median warming difference of 0.14 °C per decade (Figure 3a). This implies that across CEU heat extremes have warmed by about 50% more than the mean. On the other hand, the warming of heat extremes and mean is about the same in NEU and MED (difference in trends between TXx − TG_{JJA} not significantly different from zero, p value >0.05, Figures 3c and 3e). Hence, it is only in CEU that the warming in the extremes was larger than in the mean in this data set. This is remarkably consistent with projections of heat extremes both in many global climate models (GCMs) and RCMs that are projected to be amplified due to enhanced variability across central Europe but not necessarily over southern and northern Europe (Cattiaux et al., 2013; Fischer & Schär, 2009; Fischer et al., 2014; Orlowsky & Seneviratne, 2012; Seneviratne et al., 2006). It has been suggested that the increase in variability results from (a) land surface feedbacks, which are particularly relevant for central Europe, a transition region between a wet regime in the north and a dry regime in the south (Cattiaux et al., 2013; Fischer & Schär, 2009; Seneviratne et al., 2006), (b) the fact that the typical source regions of warming air advection in southern and continental eastern Europe warm more than the source regions of cold-air advection in northern Europe and in the Atlantic (Holmes et al., 2016), and (c) that the warming is amplified due to less clouds and higher incoming shortwave radiation (Tang et al., 2012). For future projections, Argüeso et al. (2016) find that central Europe and the Mediterranean are regions where changes in variability contribute to the increased intensity of heat waves. Therefore, it is possible that even though we do not see a significant difference between trends in summer mean temperature and extremes in the Mediterranean region so far, this could change in the future. However, Argüeso et al. (2016) also find that Europe and North America, the two regions mostly studied when looking at climate extremes are not representative for the rest of the globe, where there is less evidence for an amplified warming of heat extremes and seasonal mean warming accounts for most of the changes in heat extremes. Overall, the trends identified here are broadly consistent with earlier more regional studies focusing on daily extremes (Croitoru & Piticar, 2013; Della-Marta et al., 2007; El Kenawy et al., 2013; Fioravanti et al., 2016; Ruml et al., 2017) and hot summers (Christidis et al., 2015; Twardosz & Kossowska-Cezak, 2013).

Cold extremes have warmed across much of Europe at a rate much higher than hot extremes and with trends up to 1 °C/decade at many stations in NEU and Eastern Europe (Figure 2e). The median warming trend in cold extremes over all stations is 0.49°C per decade (Figure 2f) and thereby about 50% greater than for hot extremes. Ninety-four percent of all stations show warming trends that are significant at 40% of the stations. Since year-to-year variability in cold extremes is substantially higher than in hot extremes, it is not unexpected that the fraction of significant warming trends is smaller than for hot extremes despite the warming rate being larger. Again, the overall distribution is very different to what would be expected due to internal variability as estimated by block bootstrapping (gray distribution).

This raises the question whether cold extremes have warmed more than winter means? We find that the trends in TNn were larger than winter seasonal means in CEU and NEU (Figures 3b and 3d) but around the same in MED (Figure 3f). This implies that along with warming, winter temperature variability has decreased over CEU and NEU. Therefore, cold extremes have warmed more than the corresponding winter seasonal mean over these regions. Again, this behavior is remarkably consistent with model projections that suggest amplified warming of cold extremes over these regions (e.g., De Vries et al., 2012; Fischer et al., 2012; Holmes et al., 2016). It has been argued that the reduction in variability may relate to snow-albedo feedbacks (Fischer et al., 2011) and due to advection, that is, that the source regions of cold-air advection warm stronger than those of warm-air advection due to Arctic amplification (Screen, 2014) and land sea contrast (De Vries et al., 2012; Holmes et al., 2016). Overall cold winters in Europe are generally associated with the negative phase of the North Atlantic Oscillation (NAO)(Hurrell, 1995). Circulation patterns like the NAO as well as synoptic patterns like blocking events both contribute to thermal advection and, therefore, could play an additional role in decreased winter variability (Holmes et al., 2016), which would support stronger warming in cold extremes than winter seasonal means.

3.3. Do Regional Climate Models Reproduce Observed Trends?

In the following we address the question whether EURO-CORDEX RCMs capture the observed warming trend of hot and cold extremes. Due to data availability and to allow for a direct comparison we use gridded E-OBS instead of station data and restrict the analysis to the period 1971–2018. Even though the time

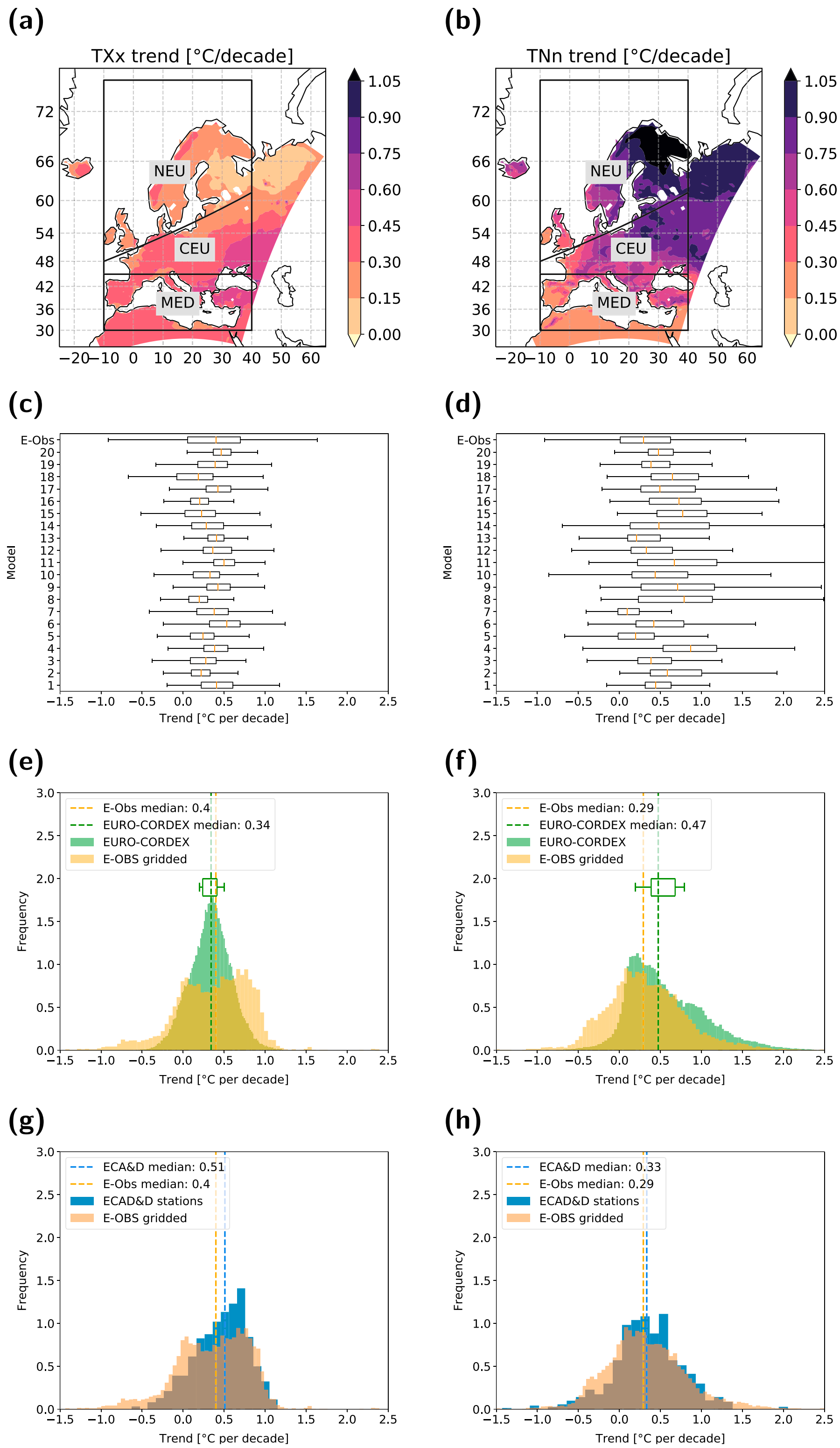

Figure 4. (a) TXx and (b) TNn multimodel mean trends in EURO-CORDEX 0.44° runs from 1971–2018 in °C/decade. (c, d) E-Obs trends and individual EURO-CORDEX modeled trends as boxplots (whiskers are 5th and 95th percentile, box 25th–75th percentile). The following panels show histograms of trends in (left) TXx and (right) TNn. Panels (e) and (f) show the gridded E-Obs data in orange versus all EURO-CORDEX 0.44° runs in green. The green box shows the medians of all individual models. Paneld (g) and (h) show the gridded E-Obs data versus the ECA&D station data in blue in °C/decade for the three SREX regions together.

period is different and based on gridded rather than on station data as in Figure 2, the pattern and range of observed trends are comparable (for the comparison gridded versus station data see Figures 4g and 4h). Trends calculated from EURO-CORDEX RCM simulations are shown in Figures 4a and 4b as multimodel means from 1971–2018 as maps and for the individual realizations in Figures 4c and 4d as boxplots. In the multimodel, which represents an estimate for the model mean forced response, the model simulations indicate larger trends in TNn than TXx. For TXx the models suggest the largest trends around the Black Sea. For TNn, trends are largest toward the North and East, potentially as result of declining snow cover and of a reduction of temperature variability due to the factors discussed above. For the individual realizations some models are closer to observations than others. Since local to regional trends particularly in extremes are strongly affected by internal variability (Fischer et al., 2014), observations should not be expected to agree in their exact pattern with the multimodel mean nor individual realizations. Thus, we aggregate data over all three subregions NEU+CEU+MED in the following.

Figures 4e and 4f show a direct comparison between the gridded E-OBS data and the EURO-CORDEX models as histograms over all models (for EURO-CORDEX) and all grid points in NEU+CEU+MED together. The gridded TXx observations show a bimodal distribution. Splitting up the data into the subregions (Figure S4) reveals that the higher end of this distribution comes from CEU and the lower end from NEU. Note that the histogram for EURO-CORDEX samples across all models, not the multimodel mean. For TXx the median trend across all EURO-CORDEX models is smaller than that of the gridded observations. Also, the observed median trend across Europe exceeds the simulated median of almost 75% of the models (boxplot in Figure 4e), consistent with earlier ENSEMBLES RCM which showed a weaker mean warming than observations (Lorenz & Jacob, 2010). Out of the 20 models more than 25% (6) simulate a roughly correct median, 25% simulate a higher median trend than E-OBS, but almost 50% (9) simulate a smaller median warming than E-OBS (Figure 4c). The differences between models and observations may partly result from unforced internal variability that affects even the distribution of multidecadal trends, from biased trends in the driving GCMs or in the RCMs or from observational uncertainties. Too little warming of hot extremes is in contrast to too much warming of hot extremes found in GCMs (Borodina et al., 2017; Fischer et al., 2014; Zwiers et al., 2011). This may relate to the fact that most of the EURO-CORDEX models used a prescribed constant aerosol climatology, while there has been a substantial decline in aerosol forcing over Europe (Wild, 2009). In addition, the missing plant response to increased CO_2 in RCMs could play a role (Giorgi & Gao, 2018; Kala et al., 2016). Also note that the observed distribution of trends is substantially wider than that sampling all simulations and all individual models (Figure 4c). This is counterintuitive and suggests that the models underestimate the spatial heterogeneity of trends at these time scales. The medians of the gridded product and the station data are similar over the combined NEU+CEU+MED region (Figure 4g).

In contrast to hot extremes, the trends in cold extremes tend to be overestimated by most of the EURO-CORDEX simulations, as the observed area median trends are lower than in 85% of the models (Figure 4d). In particular, there is hardly a model simulating a substantial fraction of negative trend, that is, the low end of the distribution is not covered by the RCMs (Figures 4d and 4f). Again, the behavior seems to be inconsistent with CMIP5 GCMs which on a global scale tend to seriously underestimate the observed warming of TNn (Fischer et al., 2014; Min et al., 2013).

4. Conclusions

We detect a clear signal from climate change in the trends in extreme temperature and heat stress based on observational data that cannot be explained by internal variability. We demonstrate that on average across Europe the number of days with extreme heat and heat stress has more than tripled from 1950–2018 from less than 2 days to more than 6 days per year. Changes are consistent across subregions, daytime and nighttime temperatures, and across different percentile thresholds. Likewise, the intensity of daily (TXx) to weekly hot extremes has increased by about 2.3 °C (median across Europe) from 1950–2018. The median rate of change of 0.33 °C per decade is larger than the global average temperature warming of about 0.2 °C per decade today (Masson-Delmotte et al., 2018). In CEU, the subregion with the strongest intensification, hot extremes have warmed about 50% more than the corresponding summer mean temperatures while in NEU and MED hot extreme and mean trends are similar.

Given that trends in annual temperature maxima experience high internal variability, it is not surprising that few stations show no trend or even a cooling. When aggregating across all of Europe a clear signal can be detected. The vast majority of 94% show a warming trend and 60% of all stations even a significantly positive trend. We demonstrate that even when accounting for spatial and temporal autocorrelation the distribution of trends across Europe cannot be explained by internal variability alone. We show for the first time for Europe that the overall signal can even be clearer detected for extreme heat stress, expressed as the annual maximum daily mean wet-bulb temperature.

At the same time extreme cold days and nights have decreased by a factor of 2–3 on average across Europe from 1950–2018. Cold extremes have warmed on average by 0.49 °C per decade, which is more than 3 °C from 1950–2018. Thereby, the warming of cold extremes in NEU and CEU is substantially higher than the corresponding winter mean warming and about 2.5 times larger than todays global average temperature warming of about 0.2 °C per decade (Masson-Delmotte et al., 2018). Again a clear signal can be detected at European scale: 94% of all stations show warming trends and 40% a statistically significant trend, an asymmetry that cannot be explained by internal variability.

The overall signal in cold and hot extremes is consistent between station network and gridded E-Obs data. Nevertheless, the exact trend magnitudes and small-scale patterns need to be interpreted with caution because not all underlying station data are complete or have been homogenized (Hofstra et al., 2009). This can result in spurious or doubtful trends at individual stations (Cornes & Jones, 2013) or over subregions. Therefore, we carefully select only stations that provide nearly complete time coverage and focus on the overall trend distributions across the whole European continent containing many stations. By focusing on the overall distribution and the median change we expect our findings to be less affected by potential inhomogeneities or biases at individual stations. By aggregating spatially we furthermore account for the fact that local trends are highly affected by internal variability.

We further demonstrate that the majority of EURO-CORDEX RCMs, which have not been evaluated regarding trends in temperature extremes, tend to underestimate the intensification of hot extremes and even more so overestimate the warming of extreme cold temperatures. This behavior is opposite to the behavior of GCMs evaluated across larger scale. We expect that the deviation of EURO-CORDEX models may partly relate to prescribing constant aerosol forcing (Giorgi & Gao, 2018), unforced internal variability that affects even the continental average trends over almost 50 years or to a biased model response to the forcing (e.g., missing plant stomatal response to increased CO_2; Kala et al., 2016). Thus, it is unclear to what extent this potential bias also affects future projections by these models.

Acknowledgments
We acknowledge the data providers in the ECA&D project. Klein Tank, A. M. G., and coauthors, 2002. Daily data set of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. of Climatol.*, 22, 1441–1453. Data and metadata available at http://www.ecad.eu. We acknowledge the E-OBS data set from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service. Cornes, R., G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones. 2018: An Ensemble Version of the E-OBS Temperature and Precipitation Datasets, *J. Geophys. Res. Atmos.*, 123. doi:10.1029/2017JD028200. E-OBS data are available at http://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php. We also acknowledge the World Climate Research Program's Working Group on Regional Climate and the Working Group on Coupled Modeling, the coordinating bodies behind CORDEX and CMIP5. We, in particular, thank the climate modeling groups (listed in Table S1 accompanying this paper) for producing and making available their model output, Urs Beyerle for downloading the data, and Sven Kotlarski for further EURO-CORDEX information. We acknowledge the Earth System Grid Federation infrastructure, an international effort led by the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison, the European Network for Earth System Modeling, and other partners in the Global Organisation for Earth System Science Portals (GO-ESSP). EURO-CORDEX data are available from ESGF, see https://euro-cordex.net/060378/index.php.en. We also thank the two anonymous reviewers who helped to improve the manuscript.

References

Alexander, L. V. (2016). Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*, *11*, 4–16. https://doi.org/10.1016/j.wace.2015.10.007

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, *111*, D05109. https://doi.org/10.1016/j.wace.2015.10.007

Argüeso, D., Di Luca, A., Perkins-Kirkpatrick, S. E., & Evans, J. P. (2016). Seasonal mean temperature changes control future heat waves. *Geophysical Research Letters*, *43*, 7653–7660. https://doi.org/10.1002/2016GL069408

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & García-Herrera, R. (2011). The hot summer of 2010: Redrawing the temperature record map of Europe. *Science (New York, N.Y.)*, *332*(6026), 220–4. https://doi.org/10.1126/science.1201224

Borodina, A., Fischer, E. M., & Knutti, R. (2017). Emergent constraints in climate projections: A case study of changes in high-latitude temperature variability. *Journal of Climate*, *30*(10), 3655–3670. https://doi.org/10.1175/JCLI-D-16-0662.1

Cattiaux, J., Douville, H., & Peings, Y. (2013). European temperatures in CMIP5: Origins of present-day biases and future uncertainties. *Climate Dynamics*, *41*(11–12), 2889–2907. https://doi.org/10.1007/s00382-013-1731-y

Cattiaux, J., Douville, H., Schoetter, R., Parey, S., & Yiou, P. (2015). Projected increase in diurnal and interdiurnal variations of European summer temperatures. *Geophysical Research Letters*, *42*, 899–907. https://doi.org/10.1007/s00382-013-1731-y

Christidis, N., Jones, G. S., & Stott, P. A. (2015). Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nature Climate Change*, *5*(1), 46–50. https://doi.org/10.1038/nclimate2468

Cornes, R. C., & Jones, P. D. (2013). How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *Journal of Geophysical Research: Atmospheres*, *118*, 10,262–10,276. https://doi.org/10.1002/jgrd.50799

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, *123*, 9391–9409. https://doi.org/10.1029/2017JD028200

Coumou, D., & Robinson, A. (2013). Historic and future increase in the global land area affected by monthly heat extremes. *Environmental Research Letters*, *8*(3), 34018. https://doi.org/10.1088/1748-9326/8/3/034018

Croitoru, A. E., & Piticar, A. (2013). Changes in daily extreme temperatures in the extra-Carpathians regions of Romania. *International Journal of Climatology*, *33*(8), 1987–2001. https://doi.org/10.1002/joc.3567

De Vries, H., Haarsma, R. J., & Hazeleger, W. (2012). Western European cold spells in current and future climate. *Geophysical Research Letters*, *39*, L04706. https://doi.org/10.1029/2011GL050665

Della-Marta, P. M., Haylock, M. R., Luterbacher, J., & Wanner, H. (2007). Doubled length of western European summer heat waves since 1880. *Journal of Geophysical Research*, *112*, D15103. https://doi.org/10.1029/2007JD008510

Diffenbaugh, N. S., Pal, J. S., Giorgi, F., & Gao, X. (2007). Heat stress intensification in the Mediterranean climate change hotspot. *Geophysical Research Letters*, *34*, L11706. https://doi.org/10.1029/2007GL030000

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., et al. (2013). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, *118*, 2098–2118. https://doi.org/10.1002/jgrd.50150

El Kenawy, A., López-Moreno, J. I., & Vicente-Serrano, S. M. (2013). Summer temperature extremes in northeastern Spain: Spatial regionalization and links to atmospheric circulation (1960–2006). *Theoretical and Applied Climatology*, *113*(3-4), 387–405. https://doi.org/10.1007/s00704-012-0797-5

Fioravanti, G., Piervitali, E., & Desiato, F. (2016). Recent changes of temperature extremes over Italy: An index-based analysis. *Theoretical and Applied Climatology*, *123*(3-4), 473–486. https://doi.org/10.1007/s00704-014-1362-1

Fischer, E. M., & Knutti, R. (2013). Robust projections of combined humidity and temperature extremes. *Nature Climate Change*, *3*(2), 126–130. https://doi.org/10.1038/nclimate1682

Fischer, E. M., & Knutti, R. (2014). Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophysical Research Letters*, *41*, 547–554. https://doi.org/10.1002/2013GL058499

Fischer, E. M., Lawrence, D. M., & Sanderson, B. M. (2011). Quantifying uncertainties in projections of extremes-a perturbed land surface parameter experiment. *Climate Dynamics*, *37*(7-8), 1381–1398. https://doi.org/10.1007/s00382-010-0915-y

Fischer, E. M., Rajczak, J., & Schär, C. (2012). Changes in European summer temperature variability revisited. *Geophysical Research Letters*, *39*, L19702. https://doi.org/10.1029/2012GL052730

Fischer, E. M., & Schär, C. (2009). Future changes in daily summer temperature variability: Driving processes and role for temperature extremes. *Climate Dynamics*, *33*(7-8), 917–935. https://doi.org/10.1007/s00382-008-0473-8

Fischer, E. M., & Schär, C. (2010). Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, *3*(6), 398–403. https://doi.org/10.1038/NGEO866

Fischer, E. M., Sedláček, J., Hawkins, E., & Knutti, R. (2014). Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, *41*, 8554–8562. https://doi.org/10.1002/2014GL062018

Frich, P., Alexander, L. V., Della-Marta, P. M., Gleason, B., Haylock, M., Klein Tank, A. M. G., & Peterson, T. (2002). Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, *19*, 193–212.

Giorgi, F., & Gao, X.-J. (2018). Regional earth system modeling: Review and future directions. *Atmospheric and Oceanic Science Letters*, *11*(2), 189–197. https://doi.org/10.1080/16742834.2018.1452520

Gross, M. H., Donat, M. G., & Alexander, L. V. (2019). Changes in daily temperature extremes relative to the mean in CMIP5 models and observations. *International Journal of Climatology*, joc.6138. https://doi.org/10.1002/joc.6138

Hofstra, N., Haylock, M., New, M., & Jones, P. D. (2009). Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research*, *114*, D21101. https://doi.org/10.1029/2009JD011799

Holmes, C. R., Woollings, T., Hawkins, E., & de Vries, H. (2016). Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *Journal of Climate*, *29*(6), 2221–2236. https://doi.org/10.1175/JCLI-D-14-00735.1

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*(5224), 676–679.

Kala, J., De Kauwe, M. G., Pitman, A. J., Medlyn, B. E., Wang, Y.-P., Lorenz, R., & Perkins-Kirkpatrick, S. E. (2016). Impact of the representation of stomatal conductance on model projections of heatwave intensity. *Scientific Reports*, *6*(1–7), 23418. https://doi.org/10.1038/srep23418

Katz, R. W., & Brown, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, *21*, 289–302.

King, A. D. (2017). Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future*, *5*, 1156–1168. https://doi.org/10.1002/2017EF000611

Kjellström, E., Bärring, L., Jacob, D., Jones, R., Lenderink, G., & Schär, C. (2007). Modelling daily temperature extremes: Recent climate and future changes over Europe. *Climatic Change*, *81*(S1), 249–265. https://doi.org/10.1007/s10584-006-9220-5

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, *22*(12), 1441–1453. https://doi.org/10.1002/joc.773

Knutson, T. R., & Ploshay, J. J. (2016). Detection of anthropogenic influence on a summertime heat stress index. *Climatic Change*, *138*(1-2), 25–39. https://doi.org/10.1007/s10584-016-1708-z

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., et al. (2014). Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, *7*(4), 1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Loikith, P. C., Neelin, J. D., Meyerson, J., & Hunter, J. S. (2018). Short warm-side temperature distribution tails drive hot spots of warm temperature extreme increases under near-future warming. *Journal of Climate*, *31*(23), 9469–9487. https://doi.org/10.1175/JCLI-D-17-0878.1

Lorenz, P., & Jacob, D. (2010). Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40. *Climate Research*, *44*(2-3), 167–177. https://doi.org/10.3354/cr00973

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., et al. (2018). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J Skea, P. R. Shukla, et al. (Eds.), *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 C* (pp. 32). Geneva, Switzerland: World Meteorological Organization. Retrieved from https://www.ipcc.ch/site/assets/uploads/sites/2/2018/07/SR15_SPM_High_Res.pdf

Mearns, L. O., Katz, R. W., & Schneider, S. H. (1984). Extreme high-temperature events: Changes in their probabilities with changes in mean temperature. *Journal of Applied Meteorology*, *23*, 1601–1613.

Min, S.-K., Zhang, X., Zwiers, F., Shiogama, H., Tung, Y.-S., & Wehner, M. (2013). Multimodel detection and attribution of extreme temperature changes. *Journal of Climate*, *26*(19), 7430–7451. https://doi.org/10.1175/JCLI-D-12-00551.1

Orlowsky, B., & Seneviratne, S. I. (2012). Global changes in extreme events: Regional and seasonal dimension. *Climatic Change*, *110*(3-4), 669–696. https://doi.org/10.1007/s10584-011-0122-9

Pal, J. S., & Eltahir, E. (2016). Future temperature in southwest Asia projected to exceed a threshold for human adaptability. *Nature Climate Change*, *6*(2), 197–200. https://doi.org/10.1038/nclimate2833

Perkins, S. E., & Fischer, E. M. (2013). The usefulness of different realizations for the model evaluation of regional trends in heat waves. *Geophysical Research Letters*, *40*, 5793–5797. https://doi.org/10.1002/2013GL057833

Ruml, M., Gregorić, E., Vujadinović, M., Radovanović, S., Matović, G., Vuković, A., et al. (2017). Observed changes of temperature extremes in Serbia over the period 1961–2010. *Atmospheric Research*, *183*, 26–41. https://doi.org/10.1016/j.atmosres.2016.08.013

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in European summer heatwaves. *Nature*, *427*(6972), 332–6. https://doi.org/10.1038/nature02300

Scherrer, S. C., Fischer, E. M., Posselt, R., Liniger, M. A., Croci-Maspoli, M., & Knutti, R. (2016). Emerging trends in heavy precipitation and hot temperature extremes in Switzerland. *Journal of Geophysical Research: Atmospheres*, *121*, 2626–2637. https://doi.org/10.1002/2015JD024634

Screen, J. A. (2014). Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nature Climate Change*, *4*(7), 577–582. https://doi.org/10.1038/nclimate2268

Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land-atmosphere coupling and climate change in Europe. *Nature*, *443*(7108), 205–209. https://doi.org/10.1038/nature05095

Seneviratne, S. I., Nicholls, N., Easterling, D. R., Goodess, C. M., Kanae, S., Kossin, J., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, et al. (Eds.), *Managing the risks of extreme events and disasters to advance climate change adaptation* (pp. 109–230). Cambridge, UK, and New York, NY, USA: Cambridge University Press.

Stull, R. (2011). Wet-Bulb temperature from relative humidity and air temperature. *Journal of Applied Meteorology and Climatology*, *50*(11), 2267–2269. https://doi.org/10.1175/JAMC-D-11-0143.1

Tang, Q., Leng, G., & Groisman, P. Y. (2012). European hot summers associated with a reduction of cloudiness. *Journal of Climate*, *25*(10), 3637–3644. https://doi.org/10..1175/JCLI-D-12-00040.1

Twardosz, R., & Kossowska-Cezak, U. (2013). Exceptionally hot summers in Central and Eastern Europe (1951–2010). *Theoretical and Applied Climatology*, *112*(3–4), 617–628. https://doi.org/10.1007/s00704-012-0757-0

Wild, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research*, *114*, D00D16. https://doi.org/10.1029/2008JD011470

Wilks, D. S. (1997). Resampling hypothesis tests for autocorrelated fields. *Journal of Climate*, *10*(1), 65–82. https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2

Wilks, D. S. (2006). On "Field Significance" and the false discovery rate. *Journal of Applied Meteorology and Climatology*, *45*, 1181–1189.

Wilks, D. S. (2016). The stippling shows statistically significant grid points. *Bulletin of the American Meteorological Society*, *97*, 2263–2274. https://doi.org/10.1175/BAMS-D-00267.1

Yiou, P., Dacunha-Castelle, D., Parey, S., & Hoang, T. T. H. (2009). Statistical representation of temperature mean and variability in Europe. *Geophysical Research Letters*, *36*, L04710. https://doi.org/10.1029/2008GL036836

Zwiers, F. W., Zhang, X., & Feng, Y. (2011). Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate*, *24*(3), 881–892. https://doi.org/10.1175/2010JCLI3908.1