AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079102

Key Points:
- Land-atmosphere coupling can amplify heat extremes under drying soil moisture
- We evaluate this coupling in CMIP5 models against flux tower observations
- We show many CMIP5 models overamplify heat extremes in wet regions by overestimating the strength of land surface coupling

Supporting Information:
- Supporting Information S1

Correspondence to:
A. M. Ukkola,
anna.ukkola@anu.edu.au

Citation:
Ukkola, A. M., Pitman, A. J., Donat, M. G., De Kauwe, M. G., & Angélil, O. (2018). Evaluating the contribution of land-atmosphere coupling to heat extremes in CMIP5 models. *Geophysical Research Letters*, *45*, 9003–9012. https://doi.org/10.1029/2018GL079102

Received 6 JUN 2018
Accepted 17 AUG 2018
Accepted article online 24 AUG 2018
Published online 5 SEP 2018

Evaluating the Contribution of Land-Atmosphere Coupling to Heat Extremes in CMIP5 Models

A. M. Ukkola[1,2], A. J. Pitman[2], M. G. Donat[3], M. G. De Kauwe[2], and O. Angélil[3]

[1]ARC Centre of Excellence for Climate Extremes, Australian National University, Canberra, ACT, Australia, [2]ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, NSW, Australia, [3]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia

Abstract Land-atmosphere coupling can amplify heat extremes under declining soil moisture. Here we evaluate this coupling in 25 Coupled Model Intercomparison Project Phase 5 models using flux tower observations over Europe and North America. We compared heat extremes (2.5% of the hottest days of the year) and the evaporative fraction (EF; a measure of land surface dryness) on the day the heat extremes occurred. We found a negative relationship between the magnitude of heat extremes and EF in both models and observations in transitional regions, with the hottest temperatures occurring during the driest days, with a similar but less certain relationship in dry regions. Surprisingly, many models also showed an amplification of heat extremes by low EF in wet regions, a finding not supported by observations. Many models may therefore overamplify heat extremes over wet regions by overestimating the strength of land-atmosphere coupling, with consequences for future projections of heat extremes.

Plain Language Summary Hot days are expected to become more frequent with climate change. Heat extremes can be exacerbated by dry conditions due to a lack of evaporative cooling. This has been shown to be an important mechanism for future amplification of heat extremes in several regions. We evaluated how well global climate models can simulate these interactions between hot and dry conditions under current climate conditions. We showed that many models overestimate this interaction in wet regions and consequently overamplify heat extremes. Our study points to an area of necessary model improvement to increase confidence in future projections of heat extremes.

1. Introduction

Heatwaves have far-reaching consequences, with harmful impacts on human health, ecosystems, infrastructure, and the economy (Perkins, 2015). Global warming is projected to exacerbate future heat extremes, with increases expected over most regions (Collins et al., 2013). Despite climate models generally agreeing on the increasing frequency of heat extremes, the magnitude of warming varies strongly across individual models (Sillmann et al., 2013). Some of the differences arise from varying levels of background warming, but past studies have shown these also relate to differences in simulated land surface feedbacks associated with heat extremes (Donat et al., 2017; Sippel et al., 2017; Vogel et al., 2017).

The land surface controls the partitioning of available energy between latent and sensible heat and can act to mitigate or amplify heat extremes (Seneviratne et al., 2010). In wet *energy-limited* conditions, when soil moisture does not limit evaporation, much of the available energy is converted into latent heat. This *evaporative cooling* reduces the sensible heat flux and limits increases in air temperature. In these conditions, latent heat and air temperature are positively related, with higher temperatures commonly occurring in association with higher latent heat fluxes. In *water-limited* conditions with drying soils, latent heat is reduced and more of the available energy is converted into sensible heat, amplifying air temperatures. In this case, latent heat and air temperature are negatively correlated, with reduced latent heat increasing temperatures. This relationship is expected to weaken in very dry environments, where soil moisture is fully depleted and no amplification of air temperatures associated with a drying soil is possible.

This conceptual understanding of the latent heat-temperature relationship has been observed during several major heatwaves (Hirschi et al., 2011; Mueller & Seneviratne, 2012; Teuling et al., 2010; Whan et al., 2015). Using flux tower observations, Teuling et al. (2010) showed that reduced evaporative cooling due to soil moisture depletion contributed to the record-breaking heatwave over Europe in 2003. Similarly, summer heat extremes were exacerbated as a consequence of reduced moisture during the 1930s U.S. Dust Bowl

10.1029/2018GL079102

drought, when a number of temperature records were set that still hold today (Donat et al., 2016). Many modeling studies have also reported a link between land surface feedbacks and heat extremes in historical and future simulations (Donat et al., 2017; Fischer, Seneviratne, Lüthi, & Schär, 2007; Fischer, Seneviratne, Vidale, et al., 2007; Seneviratne et al., 2006; Sippel et al., 2017). Donat et al. (2017) showed that CMIP5 models project greater increases in heat extremes than mean annual temperatures over Europe, North, and South America due to reduced evaporative cooling consistent with drying soils. However, they also showed that the amplified warming of heat extremes in historical simulations was inconsistent with observations (except over Europe), pointing to possible shortcomings in CMIP5 models. Similarly, Sippel et al. (2017) compared CMIP5 models to observed evapotranspiration-temperature relationships and found that several models overestimated land surface feedbacks during the summer months. By excluding models outside the observed range, they showed smaller projected increases in heat extremes in the constrained CMIP5 ensemble.

While these past studies have pointed to limitations in CMIP5 models in simulating land-atmosphere coupling on heat extremes, it remains unclear how well individual climate models simulate this coupling. We therefore evaluate the capacity of 25 CMIP5 models to simulate heat extremes as they relate to land-atmosphere coupling. Unlike previous studies that have related monthly ET or rainfall to temperature extremes (Mueller & Seneviratne, 2012; Sippel et al., 2017; Stegehuis et al., 2013), we explore the relationship between surface fluxes and heat extremes on the day the extremes occur. We compare models against flux tower data, the most direct observations of latent and sensible heat fluxes available. We evaluate model performance over Europe and North America as these regions have been identified as hot spots for future amplification of heat extremes (Seneviratne et al., 2016) and are relatively rich in flux tower observations.

We analyze the models and observations across an aridity gradient, hypothesizing that the strength of land-atmosphere coupling should vary with background climate dryness. We expect a weak control of land-atmosphere coupling on heat extremes in wetter energy-limited regions and stronger control in drier water-limited regions, with the strongest coupling in transitional zones (Koster et al., 2004). We use this framework to identify process-level limitations in individual CMIP5 models as an important step toward reducing uncertainties in future projections of heat extremes and eventually helping to improve the simulation of heat extremes in the models.

2. Materials and Methods

2.1. Definition of Hot and Dry Extremes

We used the annual 97.5th percentile value of daily maximum near-surface air temperatures (T_{max}; °C) as the threshold for heat extremes. Hot days were defined as days with T_{max} above this threshold, corresponding to approximately the nine hottest days of the year. Days below 25 °C were excluded from the analysis. Using several hot days per year allowed establishing more robust relationships and evaluation due to a larger sample size. However, as the individual hot days may not be independent, but rather part of the same heat event, we also repeated the analysis using the single hottest day of the year (commonly termed TX_x; Klein Tank et al., 2009) but found the results to be qualitatively consistent (not shown).

We used the evaporative fraction (EF; unitless) to characterize land surface dryness. It is calculated from latent (Q_{le}; W/m^2) and sensible (Q_h; W/m^2) heat as

$$\mathrm{EF} = Q_{le}/(Q_{le} + Q_h) \tag{1}$$

EF describes the partitioning of available energy between latent and sensible heat, with low values (EF → 0) characterizing dry conditions and high values (EF → 1) wet conditions. We extracted daily mean Q_{le} and Q_h and then derived EF on the specific days the heat extremes occurred. Occasional EF values below 0 and above 1 were excluded from the analysis.

We determined the coupling strength between T_{max} and EF using linear regression slopes ($\Omega_{EF,Tmax}$; °C):

$$\Omega_{EF,Tmax} = \Delta T_{max}/\Delta \mathrm{EF} \tag{2}$$

We calculated the slopes using linear quantile regression (Koenker, 2005) for quantiles ranging from 0.05 to 0.95. This was done to explore coupling behavior across the whole data distribution as observed coupling strength has been shown to vary with the magnitude of temperature extremes (Hirschi et al., 2011). Quantile regression also avoids the parametric assumptions underlying traditional linear regression methods. This method is similar to those used in previous studies investigating land surface coupling during heat extremes (Hirschi et al., 2011; Lorenz et al., 2015; Seneviratne et al., 2006). While some studies have relied on correlation, we used the regression slope as it indicates both the direction and magnitude of coupling. We calculated the quantile regression slopes and their associated uncertainty using R (package quantreg).

2.2. Models and Data

We used half-hourly observations of Q_{le}, Q_{h}, and T_{max} from 48 flux towers over Europe and North America (supporting information Figure S1). The data were obtained from the November 2016 FLUXNET2015 and La Thuile 2007 archives (http://fluxnet.fluxdata.org/). Only open access Tier 1 sites from FLUXNET2015 and fair use sites from La Thuile were used. Twenty-four towers were located in North America and 24 towers in Europe.

The data were first gap filled using the FluxnetLSM R package (Ukkola et al., 2017). We then selected sites and years that had $<$10% missing or gap-filled half-hourly time steps for all three data variables during the May–September warm season. The total data period spans from 1995 to 2014, but the available years vary by site. The average number of site years was 8.2, ranging from 1 to 19 years.

For CMIP5, we used daily outputs of T_{max}, Q_{le} and Q_{h} from 25 models (variables *tasmax*, *hfls*, and *hfss* in the CMIP5 archive, respectively; supporting information Table S1). We analyzed model outputs during the period 1990–2010, which broadly corresponds to the period of the flux tower observations. We combined simulations from the historical experiment during 1990–2005 and the Representative Concentration Pathway 8.5 (RCP8.5) experiment during 2006–2010. We used one ensemble member per model to avoid biasing the results toward models with more available members. The simulation r6i1p1 was used for CCSM4 and r1i1p1 for all other models. We masked out all grid cells that had a land fraction ≤90%. We also used outputs for tree fraction (CMIP5 variable *treeFrac*) to investigate model behavior due to land cover type. For models with a time-varying land cover, the average tree fraction during 1990–2005 was used.

2.3. Aridity Classification

We used the aridity index AI to characterize climatological dryness (AI = PET/*P*, where PET is mean annual potential evapotranspiration and *P* is the mean annual precipitation). For CMIP5 models, AI was calculated separately for each model and grid cell using the 1990–2010 climatology for annual P and PET. PET was calculated following Milly and Dunne (2016) as

$$\mathrm{PET} = 0.8\ (Q_{\mathrm{le}} + Q_{\mathrm{h}}) \tag{3}$$

For observations, two estimates of AI were calculated for each flux tower site. The first estimate was calculated from site observations of annual P and PET (as per equation (3)) for years with good quality data (see section 2.2). This offers the most direct estimate of AI for each flux tower site but does not necessarily reflect long-term climatological AI for sites with a limited number of data years. As such, we also calculated a second estimate of AI from monthly gridded P and PET from the Climatic Research Unit (CRU) TS 3.23 data product (Harris et al., 2014) during 1990–2010. As the surface fluxes required to calculate PET using equation (3) were not available, we approximated CRU-based PET from equilibrium evapotranspiration (ET_0; mm/day) following Priestley and Taylor (1972):

$$\mathrm{ET}_0 = \Delta/(\Delta + \gamma)\ R_n \tag{4}$$

where Δ is the rate of increase of saturated vapor pressure with temperature (Pa/K), γ the psychrometric constant (65 Pa/K), and R_n is net radiation (W/m^2). ET_0 was calculated from CRU mean monthly air temperature and cloud cover using the R package SPLASH (Davis et al., 2017). Data for the pixel incorporating each tower were used to calculate AI. We show results using the direct site estimates in the main manuscript and those using the CRU-based AI in the supporting information.

We classified models and observations into discrete aridity categories: *wet* (AI $<$ 1), *transitional* (1 ≤ AI $<$ 1.54), and *dry* (AI ≥ 1.54) (United Nations Environment Programme, 1997). Note that our definition of transitional is lower than in some previous studies (e.g., Greve et al., 2014). We selected these thresholds to reflect the

10.1029/2018GL079102

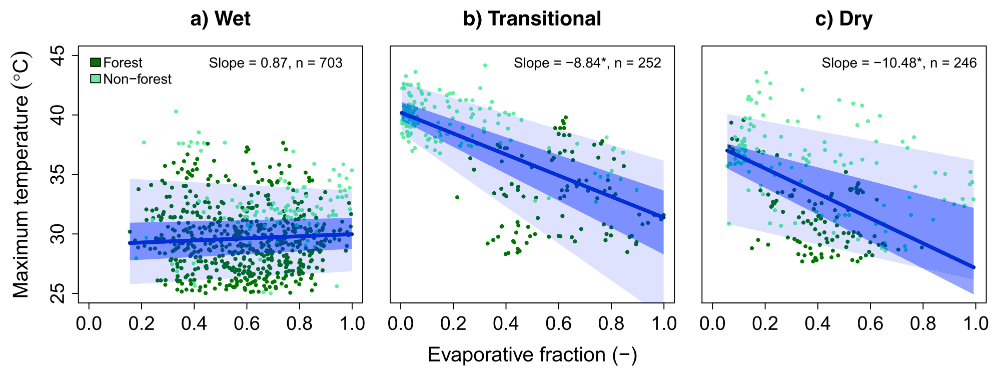

Figure 1. The relationship between observed evaporative fraction and maximum temperature at the flux tower sites on the day of heat extremes, classified by the aridity index. The solid line shows the median quantile regression slope. The dark shading indicates the range of the 0.3 and 0.7 quantile regression slopes and the light shading the range of the 0.1 and 0.9 quantile regression slopes. The median slope value is shown for each figure, with an asterisk indicating a statistically significant ($p < 0.05$) slope. European and North American sites are shown together. Site-based aridity index was used to classify sites here; an equivalent plot using Climatic Research Unit-based aridity index is shown in Figure S4. Sites with broadleaf or needleleaf vegetation type were classified as forests and all other sites were classified as non-forest. There were 26, 8, and 13 sites in the wet, transitional, and dry categories, respectively.

aridity of the study areas (limited extent of semiarid to arid regions) and to ensure sufficient data coverage in each category. The AI masks for each model are shown in Figures S2 and S3. We calculated $\Omega_{EF,T\max}$ separately for the three aridity classes using all pixels or sites across Europe and North America that fell into each class.

3. Results and Discussion

3.1. Coupling Between Hot and Dry Extremes

We first explore the EF-T_{max} relationship in the flux tower observations for three aridity categories (Figure 1). The observations are consistent with our hypothesis of a weak EF-T_{max} relationship in the wet regions (median $\Omega_{EF,T\max} = 0.87$, p value $p > 0.1$). In contrast, strong negative coupling between EF and T_{max} is observed in the transitional (median $\Omega_{EF,T\max} = -8.84$, $p < 0.001$) and dry regions (median $\Omega_{EF,T\max} = -10.48$, $p < 0.001$). This suggests that heat extremes are amplified by dry conditions in these regions, with the hottest maximum temperatures on average coinciding with the driest conditions.

The observed relationship is strongest at the arid sites, but median $\Omega_{EF,T\max}$ values are smaller when calculated across individual sites (Figure S5c). This may suggest that $\Omega_{EF,T\max}$ is overestimated in the arid zones, possibly due to the $\Omega_{EF,T\max}$ fitted to all sites reflecting a latitudinal temperature gradient. Meanwhile, the overall $\Omega_{EF,T\max}$ is slightly weaker in the transitional sites but the majority of these sites show negative slopes in line with the overall response across sites (Figure S5b). Land cover explains some of the variability across sites, with forested sites generally showing a weaker coupling, as expected associated with greater access to soil water and consequently a higher Q_{le}. However, this behavior is partly driven by one anomalous site (US-UMB) with a near-zero $\Omega_{EF,T\max}$ (Figure S5b). Overall, this suggests a consistent negative coupling between EF and T_{max} in transitional regions, in line with expectation. Similarly, the majority of wet sites show weak relationships, in agreement with the relationship fitted to all sites (Figure S5a). While several individual sites show opposing signs, only two wet sites have a significant negative $\Omega_{EF,T\max}$ and the weak coupling was consistent across forested and nonforested sites.

We next evaluate the 25 CMIP5 models against this observed behavior. Figure 2 shows the median $\Omega_{EF,T\max}$ for each model and for observations for the three aridity classes (the underlying data for each model are shown in Figures S6–S8). In the transitional and dry regions of Europe and North America, all the models simulate a negative $\Omega_{EF,T\max}$ (Figures 2b and 2c), suggesting an amplification of heat extremes by low water availability in line with the observations (Figures 1b and 1c). Eight out of 25 models fall within the observed magnitude of $\Omega_{EF,\mathrm{Tmax}}$ in the transitional regions, with other models simulating a weaker

10.1029/2018GL079102

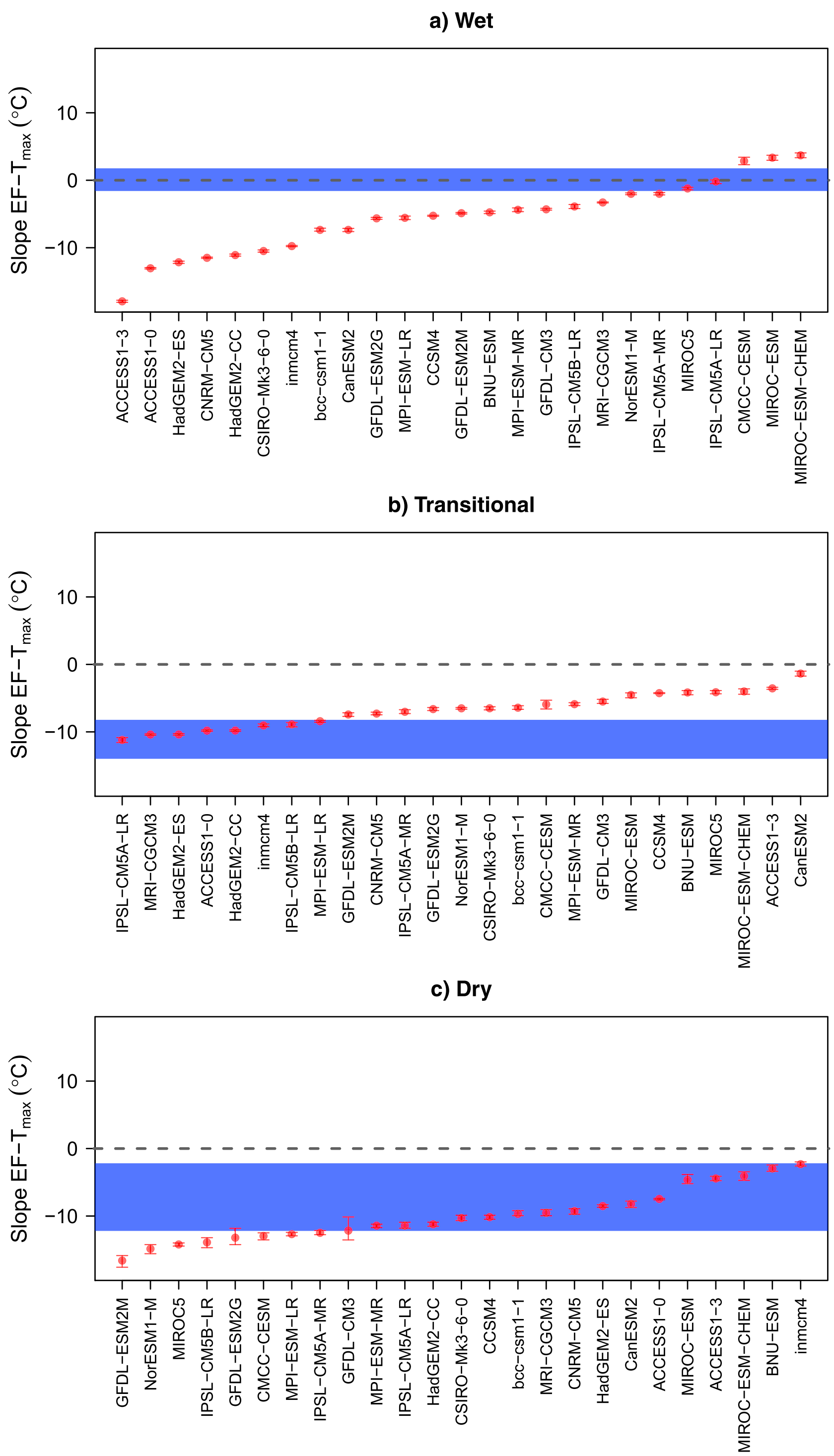

Figure 2. The relationship between evaporative fraction and maximum temperature for models and observations classified by the aridity index. The median quantile regression slope between evaporative fraction (EF) and T_{max} ($\Omega_{EF,Tmax}$) is shown for each model and for observations for the three aridity classes. The red dots represent each model's slope, with error bars indicating the 95% confidence intervals. The blue shading shows the full range of the two observed slopes calculated from site- and Climatic Research Unit-based aridity, including their 95% confidence intervals. The dotted line shows the zero line.

relationship. By contrast, in the dry regions, 6 out of 25 models simulated a stronger negative $\Omega_{EF,T\max}$ than was observed, with the rest lying within the observed range. These differences in magnitude may be partly due to a mismatch in the spatial extent between site-scale flux tower observations (~1 km^2) and the models (>7,000 km^2), as well as incomplete spatial coverage of flux tower observations. Nevertheless, all the models agree on the negative sign of the relationship with the observations in the transitional and dry regions.

By contrast, in the wet regions, the models differ strongly showing both significant positive and negative $\Omega_{EF,T\max}$ (Figure 2a). Only 2 of 25 models lie within the observed range. Three of 25 models exhibit positive $\Omega_{EF,T\max}$ (suggesting high temperatures increase EF rather than vice versa), in line with theory. However, 20 of 25 models exhibit negative $\Omega_{EF,T\max}$ below the observed range, suggesting similar behavior to transient and dry regions whereby heat extremes are amplified due to the repartitioning of surface fluxes. This suggests that these CMIP5 models are overestimating the land-atmosphere coupling over wet regions, and consequently overamplifying heat extremes. These findings agree with Ferguson et al. (2012) who showed that offline land surface models overestimated coupling in both the wettest and driest climate regimes compared to remote sensing products and flux tower observations.

These results are consistent for linear regression slopes calculated across different quantiles (supporting information Figures S9–S11). In the wet regions, the models with the most negative $\Omega_{EF,T\max}$, including ACCESS and Hadley Centre models, consistently lie below the observed range in wet regions for all quantiles (Figure S9). In the dry and transitional regions, the models broadly capture observations at all quantiles and show little sensitivity to the choice of quantile (Figure S10 and 11), suggesting robustness of our findings across the data distribution.

3.2. Sources of Model Discrepancies

The model behavior in wet regions is consistent with biases identified in offline land surface models (LSMs). Several studies have shown that LSMs do not partition energy fluxes correctly during water-stressed periods when compared to flux tower observations but rather systematically underestimate latent heat fluxes and overestimate sensible heat fluxes during dry periods (Blyth et al., 2010; De Kauwe et al., 2015; Keenan et al., 2009; Li et al., 2012; Morales et al., 2005; Ukkola, De Kauwe, et al., 2016; Ukkola, Pitman, et al., 2016). Similar biases have been noted in CMIP5 models during water-stressed periods when compared to a global gridded ET product (Mueller & Seneviratne, 2014). While the LSMs employed inside coupled climate models can differ from their offline counterparts in process representations and parameters, it is likely that they exhibit similar behavior identified in the offline models.

For example, CABLE, SURFEX (ISBA) and JULES (used by the ACCESS, CNRM-CM5 and HadGEM2 models, respectively) strongly overestimated the occurrence of low summer-time Q_{le} compared to observations at six flux tower sites (Ukkola, De Kauwe, et al., 2016). Here the ACCESS, CNRM-CM5, and HadGEM2 models have the strongest amplification of heat extremes with EF in the wet regions (median $\Omega_{EF,T\max}$ ranging from −11.1 to −17.9; Figure 2a). In contrast, Ukkola, De Kauwe, et al., 2016 found the ORCHIDEE model (used by the IPSL models) to better capture flux tower observations, despite also having an overall tendency to underestimate summer-time Q_{le}. Here we similarly found the IPSL models to better represent the observed behavior in the wet regions (with $\Omega_{EF,T\max}$ ranging from −3.9 to −0.2). These findings also agree with Gevaert et al. (2018) who showed that JULES and SURFEX (ISBA) have stronger ET-temperature coupling in global offline simulations during the warm season compared to ORCHIDEE.

However, the intermodel differences in $\Omega_{EF,T\max}$ could not be merely explained by absolute differences in EF or Q_{le} on the day of the heat extremes. The strongest $T_{\max}$ amplification could be expected to be associated with the driest conditions, but we found no statistically significant relationship between $\Omega_{EF,T\max}$ and the absolute EF or Q_{le} values on the day of the heat extremes across any of the three aridity categories. By contrast, the ACCESS and HadGEM2 models simulated a mean EF between 0.65 and 0.69 on the days hot extremes occurred when averaged across the wet regions, slightly above the multmodel mean of 0.63. Similarly, these models simulate absolute Q_{le} values above the multimodel mean (104 to 108 W/m^2, compared to 95 W/m^2 averaged across all models). As such, the models with the most negative $\Omega_{EF,T\max}$ do not have a tendency to simulate the driest conditions in absolute terms on the days the heat extremes occur.

The models with the strongest amplification of heat extremes also tend to simulate above-average seasonal EF and Q_{le} during the May–September warm season in the wet regions. The ACCESS and HadGEM2 models simulated warm-season EF between 0.67 and 0.72, compared to the multimodel average of 0.63. Similarly, they simulated above-average Q_{le} (89 to 97 W/m^2, compared to 80 W/m^2 across all models). It should be noted the absolute differences in EF and Q_{le} can be influenced by the different spatial extent of wet regions in the individual models as these were determined using each model's long-term AI. Nevertheless, our results suggest that the stronger amplification of T_{max} in these models cannot be explained by a tendency to simulate drier conditions in absolute terms during the warm season.

The absolute differences in EF and Q_{le} provide no obvious reasons for model behavior. However, we note the seasonal distribution of EF and Q_{le} varies significantly across the models. Models with the most negative $\Omega_{EF,T\max}$ in the wet regions tend to simulate maximum daily EF and Q_{le} earlier in the warm season (Figure 3). The two exceptions are CanESM2 and BNU-ESM, which simulate peak daily Q_{le} near the end of the warm season. Excluding these two models, there is a significant correlation between the models' median $\Omega_{EF,T\max}$ and the day of peak Q_{le} (correlation coefficient $r = 0.49$, $p < 0.05$). The five models with the most negative $\Omega_{EF,T\max}$ (ACCESS, Hadley Center and CNRM models) simulate peak daily EF 23 days and peak daily Q_{le} 17 days before observed. In contrast, the five models with the most positive $\Omega_{EF,T\max}$ (IPSL-CM5A-LR, CMCC-CESM, and MIROC models) simulate peak daily EF 37 days after and peak Q_{le} 1 day before observed. This suggests that the models that most strongly amplify T_{max} with EF in the wet regions overestimate early season EF and Q_{le} relative to their overall warm season climatology. This is consistent with biases identified in several offline LSMs, with a tendency for the models to overevaporate in early warm season and underevaporate in late warm season (De Kauwe et al., 2017; Decker et al., 2017). This overevaporation causes a depletion of soil moisture stores earlier in the season, leading to water-stressed conditions and consequently lower EF during the later (and hotter) parts of the warm season. We could not explicitly evaluate the seasonal soil moisture dynamics as the models do not report daily total column soil moisture but EF should closely reflect variability in soil moisture.

One reason for differences in early season Q_{le} could arise from different leaf green-up phases, with models with higher spring leaf area index (LAI) reaching peak Q_{le} earlier in the season. We examined this for 19 models that provided LAI outputs but found no relationship between the models' early season (May–June) LAI and the timing of peak Q_{le} in the wet regions. On the contrary, the models reaching peak Q_{le} earlier than observed simulated slightly lower LAI in May and June (2.7 and 2.8 on average, respectively), compared to the models peaking after observed (3.3 both months). Similarly, these models reach maximum monthly LAI on average 0.6 months later than models with Q_{le} peaking later than observed. However, this conclusion should be interpreted tentatively as LAI was only available at monthly time steps from the CMIP5 archive. It is likely that there are compensating biases among the models related to models that overevaporate in spring and underestimated LAI magnitude and/or an incorrect seasonality and vice versa.

Model differences in $\Omega_{EF,T\max}$ could be related to land cover, with a higher tree cover expected to lead to weaker coupling due to higher Q_{le}. Land cover was found to impact the strength of coupling within individual models (Figure S12). For example, in ACCESS1-0 and the Hadley Center models, the coupling is weaker at higher tree fraction (but for Hadley models $\Omega_{EF,T\max}$ remains negative even at high tree fraction). ACCESSS1-3 shows strong coupling independent of tree fraction; this lack of sensitivity to land cover is likely due to the model using fixed root fractions across all vegetation types. Overall, the differences in $\Omega_{EF,T\max}$ across the models did not correlate with their mean tree fraction ($r = 0.04$, $p > 0.10$).

Overall, our results suggest that an overestimation of land-surface coupling is leading to an overamplification of heat extremes in many coupled climate models in the wet regions. Clearly, there is a need to reexamine process representations in the land surface model components to improve the partitioning of surface fluxes at daily to seasonal scales. Our study is limited to Europe and North America due to the sparse coverage of flux tower observations elsewhere. Moreover, we acknowledge that the differences in spatial resolution between the CMIP5 models and flux tower observations, and the uneven distribution of flux tower sites, are an unavoidable limitation in our analysis. Nevertheless, by comparing the relationship between hot and dry extremes, rather than absolute quantities, we have been able to use flux tower observations to identify process-level biases in the CMIP5 models. Our findings will help guide future development of coupled climate models and their components to reduce uncertainties in future projections of heat

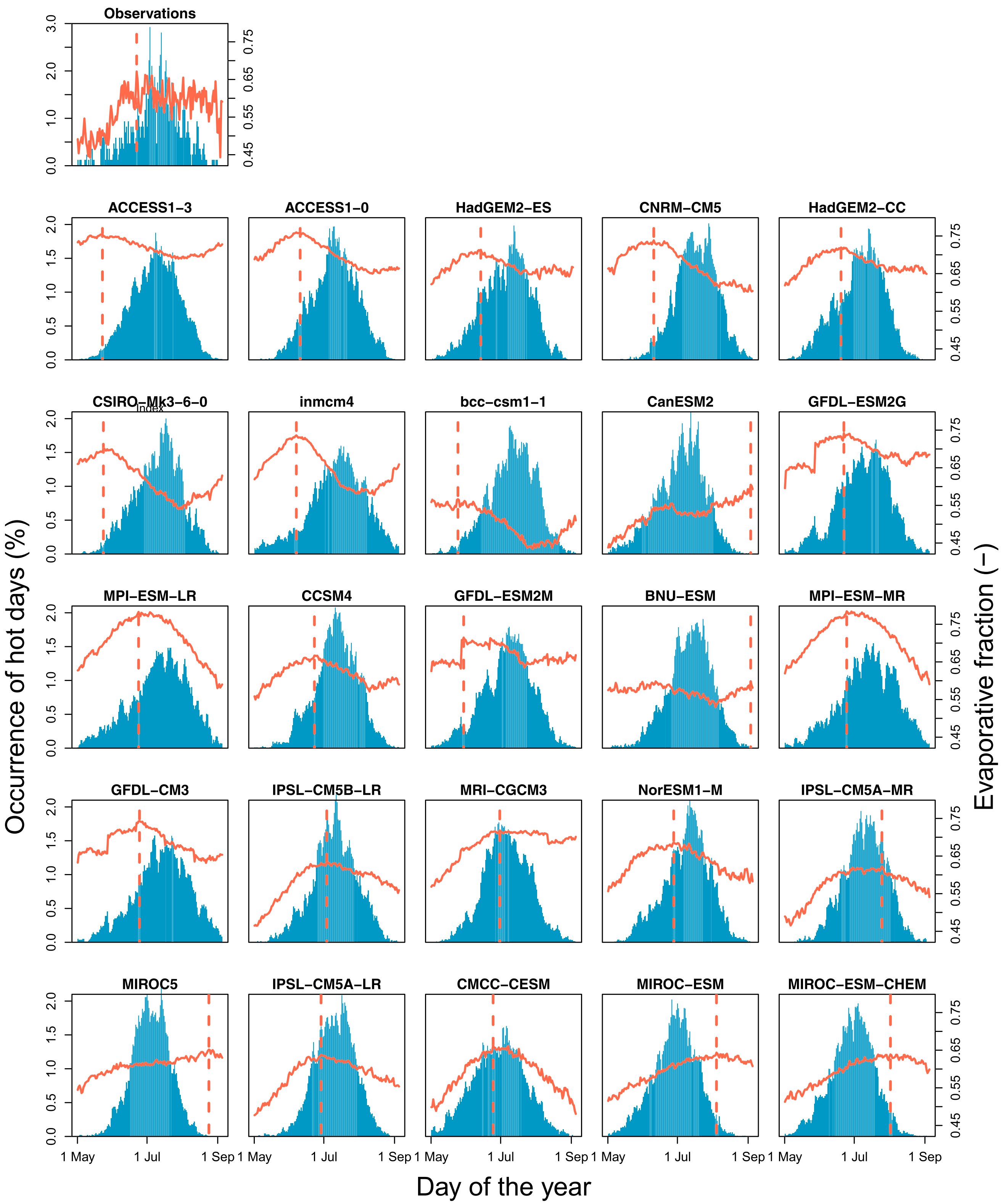

Figure 3. Daily climatology of evaporative fraction (EF) and hot day occurrence during the May–September warm season in models and observations in the wet regions. Mean daily EF is shown in red, with the day of maximum EF indicated by the vertical red line. The blue bars show the occurrence of hot days on each day (expressed as percent of total hot days). The models were arranged by their median $\Omega_{EF,Tmax}$, running from the most negative (ACCESS1-3) to positive (MIROC-ESM-CHEM). All flux towers in the wet category were combined to calculate the observed climatology (using site-based AI) and all grid cells in the wet category over Europe and America for the models.

extremes. In particular, future works focused on the representation of water stress (Bonan et al., 2014; De Kauwe et al., 2015; Egea et al., 2011; Sperry et al., 2017), LAI phenology (Hufkens et al., 2016; Lorenz et al., 2013), and the representation of soil evaporation and hydraulics (Decker et al., 2017; Jefferson & Maxwell, 2015; Lawrence & Chase, 2009) are likely to be key.

4. Conclusions

We have evaluated CMIP5 models against flux tower observations for simulating the impact of land-surface coupling to heat extremes. We showed that the models broadly agree with observations in transitional and dry regions of Europe and North America, simulating an amplification of heat extremes with dry conditions. By contrast, the models differ strongly in the wet regions, with 20 of the 25 models overamplifying hot extremes with dry conditions compared to the observations. This suggests that many CMIP5 models may be overestimating heat extremes over wet regions. Model biases are shown to relate to the seasonal distribution of latent heat and are consistent with latent heat biases identified in offline land surface models. By building on process-level understanding of heat extremes, our study contributes to an improved understanding of the sources of model differences in projections of heat extremes.

Acknowledgments
This work was supported by the ARC Centre of Excellence for Climate Extremes (CE170100023). We acknowledge the National Computational Infrastructure at the Australian National University and the Earth System Grid Federation for making the CMIP5 model outputs available. We also gratefully acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and thank the climate modeling groups (listed in Table S1) for producing and making available their model output. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

References

Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P., & Shuttleworth, J. (2010). Evaluating the JULES land surface model energy fluxes using FLUXNET data. *Journal of Hydrometeorology*, *11*(2), 509–519. https://doi.org/10.1175/2009JHM1183.1

Bonan, G. B., Williams, M., Fisher, R. A., & Olseon, K. W. (2014). Modeling stomatal conductance in the earth system: Linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geoscientific Model Development*, *7*(5), 2193–2222. https://doi.org/10.5194/gmd-7-2193-2014

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis*, *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). Cambridge, United Kingdom and New York, NY: Cambridge University Press.

Davis, T. W., Prentice, I. C., Stocker, B. D., Thomas, R. T., Whitley, R. J., Wang, H., et al. (2017). Simple process-led algorithms for simulating habitats (SPLASH v.1.0): Robust indices of radiation, evapotranspiration and plant-available moisture. *Geoscientific Model Development*, *10*(2), 689–708. https://doi.org/10.5194/gmd-10-689-2017

De Kauwe, M. G., Zhou, S.-X., Medlyn, B. E., Pitman, A. J., Wang, Y. P., Duursma, R. A., & Prentice, I. C. (2015). Do land surface models need to include differential plant species responses to drought? Examining model predictions across a latitudinal gradient in Europe. *Biogeosciences*, *12*(15), 12,349–12,393. https://doi.org/10.5194/bgd-12-12349-2015

De Kauwe, M. G., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., et al. (2017). Challenging terrestrial biosphere models with data from the long-term multifactor prairie heating and CO_2 enrichment experiment. *Global Change Biology*, *23*(9), 3623–3645. https://doi.org/10.1111/gcb.13643

Decker, M., Or, D., Pitman, A., & Ukkola, A. (2017). New turbulent resistance parameterization for soil evaporation based on a pore-scale model: Impact on surface fluxes in CABLE. *Journal of Advances in Modeling Earth Systems*, *9*(1), 220–238. https://doi.org/10.1002/2016MS000832

Donat, M. G., King, A. D., Overpeck, J. T., Alexander, L. V., Durre, I., & Karoly, D. J. (2016). Extraordinary heat during the 1930s US dust bowl and associated large-scale conditions. *Climate Dynamics*, *46*(1-2), 413–426. https://doi.org/10.1007/s00382-015-2590-5

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

Egea, G., Verhoef, A., & Vidale, P. L. (2011). Towards an improved and more flexible representation of water stress in coupled photosynthesis-stomatal conductance models. *Agricultural and Forest Meteorology*, *151*(10), 1370–1384. https://doi.org/10.1016/j.agrformet.2011.05.019

Ferguson, C. R., Wood, E. F., & Vinukollu, R. K. (2012). A global Intercomparison of modeled and observed land–atmosphere coupling*. *Journal of Hydrometeorology*, *13*(3), 749–784. https://doi.org/10.1175/JHM-D-11-0119.1

Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C. (2007). Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, *34*, L06707. https://doi.org/10.1029/2006GL029068

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., & Schär, C. (2007). Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, *20*(20), 5081–5099. https://doi.org/10.1175/JCLI4288.1

Gevaert, A. I., Miralles, D. G., de Jeu, R. A. M., Schellekens, J., & Dolman, A. J. (2018). Soil moisture-temperature coupling in a set of land surface models. *Journal of Geophysical Research: Atmospheres*, *123*, 1481–1498. https://doi.org/10.1002/2017JD027346

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, *7*(10), 716–721. https://doi.org/10.1038/ngeo2247

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., et al. (2011). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience*, *4*(1), 17–21. https://doi.org/10.1038/ngeo1032

Hufkens, K., Keenan, T. F., Flanagan, L. B., Scott, R. L., Bernacchi, C. J., Joo, E., et al. (2016). Productivity of north American grasslands is increased under future climate scenarios despite rising aridity. *Nature Climate Change*, *6*(7), 710–714. https://doi.org/10.1038/NCLIMATE2942

Jefferson, L., & Maxwell, R. M. (2015). Evaluation of simple to complex parameterizations of bare ground evaporation. *Journal of Advances in Modeling Earth Systems*, *7*(3), 1075–1092. https://doi.org/10.1002/2014MS000398

Keenan, T., García, R., Friend, A. D., Zaehle, S., Gracia, C., & Sabate, S. (2009). Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO_2 and water fluxes through combined in situ measurements and ecosystem modelling. *Biogeosciences*, *6*(1), 2285–2329. https://doi.org/10.5194/bgd-6-2285-2009

Klein Tank, A. M. G., Zwiers, F. W., & Zhang, X. (2009). *Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation*. Geneva, Switzerland: WMO.

Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.

Koster, R. D.Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Lawrence, P. J., & Chase, T. N. (2009). Climate impacts of making evapotranspiration in the community land model (CLM3) consistent with the simple biosphere model (SiB). *Journal of Hydrometeorology*, *10*(2), 374–394. https://doi.org/10.1175/2008JHM987.1

Li, L., Wang, Y.-P., Yu, Q., Pak, B., Eamus, D., Yan, J., et al. (2012). Improving the responses of the Australian community land surface model (CABLE) to seasonal drought. *Journal of Geophysical Research*, *117*, G04002. https://doi.org/10.1029/2012JG002038

Lorenz, R., Davin, E. L., Lawrence, D. M., Stöckli, R., & Seneviratne, S. I. (2013). How important is vegetation phenology for European climate and heat waves? *Journal of Climate*, *26*(24), 10,077–10,100. https://doi.org/10.1175/JCLI-D-13-00040.1

Lorenz, R., Pitman, A. J., Hirsch, A. L., & Srbinovsky, J. (2015). Intraseasonal versus interannual measures of land–atmosphere coupling strength in a global climate model: GLACE-1 versus GLACE-CMIP5 experiments in ACCESS1.3b. *Journal of Hydrometeorology*, *16*(5), 2276–2295. https://doi.org/10.1175/JHM-D-14-0206.1

Milly, P. C. D., & Dunne, K. A. (2016). Potential evapotranspiration and continental drying. *Nature Climate Change*, *6*(10), 946–949. https://doi.org/10.1038/NCLIMATE3046

Morales, P., Sykes, M. T., Prentice, I. C., Smith, P., Smith, B., Bugmann, H., et al. (2005). Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. *Global Change Biology*, *11*(12), 2211–2233. https://doi.org/10.1111/j.1365-2486.2005.01036.x

Mueller, B., & Seneviratne, S. I. (2012). Hot days induced by precipitation deficits at the global scale. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12,398–12,403. https://doi.org/10.1073/pnas.1204330109

Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, *41*, 128–134. https://doi.org/10.1002/2013GL058055

Perkins, S. E. (2015). A review on the scientific understanding of heatwaves — Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164-165*, 242–267. https://doi.org/10.1016/j.atmosres.2015.05.014

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3-4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., & Wilby, R. L. (2016). Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature*, *529*(7587), 477–483. https://doi.org/10.1038/nature16542

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

Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, *118*, 2473–2493. https://doi.org/10.1002/jgrd.50188

Sippel, S., Zscheischler, J., Mahecha, M. D., Orth, R., Reichstein, M., Vogel, M., & Seneviratne, S. I. (2017). Refining multi-model projections of temperature extremes by evaluation against land-atmosphere coupling diagnostics. *Earth System Dynamics*, *8*(2), 387–403. https://doi.org/10.5194/esd-8-387-2017

Sperry, J. S., Venturas, M. D., Anderegg, W. R. L., Mencuccini, M., Mackay, D. S., Wang, Y., & Love, D. M. (2017). Predicting stomatal responses to the environment from the optimization of photosynthetic gain and hydraulic cost. *Plant, Cell & Environment*, *40*(6), 816–830. https://doi.org/10.1111/pce.12852

Stegehuis, A. I., Teuling, A. J., Ciais, P., Vautard, R., & Jung, M. (2013). Future European temperature change uncertainties reduced by using land heat flux observations. *Geophysical Research Letters*, *40*, 2242–2245. https://doi.org/10.1002/grl.50404

Teuling, A. J., Seneviratne, S. I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., et al. (2010). Contrasting response of European forest and grassland energy exchange to heatwaves. *Nature Geoscience*, *3*(10), 722–727. https://doi.org/10.1038/ngeo950

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., et al. (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environmental Research Letters*, *11*(10), 104012. https://doi.org/10.1088/1748-9326/11/10/104012

Ukkola, A. M., Haughton, N., De Kauwe, M. G., Abramowitz, G., & Pitman, A. J. (2017). FluxnetLSM R package (v1.0): A community tool for processing FLUXNET data for use in land surface modelling. *Geoscientific Model Development Discussion*, 1–21. https://doi.org/10.5194/gmd-2017-58

Ukkola, A. M., Pitman, A. J., Decker, M., De Kauwe, M. G., Abramowitz, G., Kala, J., & Wang, Y.-P. (2016). Modelling evapotranspiration during precipitation deficits: Identifying critical processes in a land surface model. *Hydrology and Earth System Sciences*, *12*(10), 10,789–10,825. https://doi.org/10.5194/hessd-12-10789-2015

United Nations Environment Programme (1997). World Atlas of desertification.

Vogel, M. M., Orth, R., Cheruy, F., Hagemann, S., Lorenz, R., van den Hurk, B. J. J. M., & Seneviratne, S. I. (2017). Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. *Geophysical Research Letters*, *44*, 1511–1519. https://doi.org/10.1002/2016GL071235

Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O., & Seneviratne, S. I. (2015). Impact of soil moisture on extreme maximum temperatures in Europe. *Weather and Climate Extremes*, *9*, 57–67. https://doi.org/10.1016/j.wace.2015.05.001