The DOI for this manuscript is doi: 10.1175/JCLI-D-16-0491.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Herrera-Estrada, J., and J. Sheffield, 2017: Uncertainties in Future Projections of Summer Droughts and Heat Waves over the Contiguous United States. J. Climate. doi:10.1175/JCLI-D-16-0491.1, in press.

Uncertainties in Future Projections of Summer Droughts and Heat Waves over the Contiguous United States

Julio E. Herrera-Estrada[1] and Justin Sheffield[1, 2]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

[2]Geography and Environment, University of Southampton, Southampton, United Kingdom

Corresponding author address: Julio Herrera-Estrada, Princeton University, E-209A Engineering Quadrangle, Princeton, NJ 08544.

E-mail: jherrera@princeton.edu

ABSTRACT

Droughts and heat waves have important impacts on multiple sectors including water resources, agriculture, electricity generation, and public health, so it is important to understand how they will be affected by climate change. However, there is large uncertainty in the projected changes of these extreme events from climate models. We compare historical biases in models against their future projections to understand and attempt to constrain these uncertainties. Historical biases in precipitation, near-surface air temperature, evapotranspiration, and a land-atmospheric coupling metric are calculated for 24 models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) against the North American Land Data Assimilation System Phase 2 (NLDAS-2) as reference for 1979-2005. Biases are highly correlated across variables, with some models being hotter and drier, and others wetter and cooler. Models that overestimate summer precipitation project larger increases in precipitation, evapotranspiration, and land-atmospheric coupling over important agricultural regions by the end of the 21st century (2070-2099) under RCP8.5, although the percentage variance explained is low. Changes in the characteristics of droughts and heat waves are calculated and linked to historical biases in precipitation and temperature. A method to constrain uncertainty by ranking models based on historical performance is discussed but the rankings differ widely depending on the variable considered. Despite the large uncertainty that remains in the magnitude of the changes, there is consensus amongst models that droughts and heat waves will increase in multiple regions in the US by the end of the 21st century unless climate mitigation actions are taken.

1. INTRODUCTION

Droughts and heat waves are two of the most damaging natural hazards that affect water resources (Dawadi et al. 2012), agriculture (Lesk et al. 2016), electricity generation (Vliet et al. 2016), and public health (Anderson and Bell 2011). When these extreme events impact large expanses of cultivated areas, they can cause water and heat stress to plants and crops (Lobell et al. 2013; Hatfield and Prueger 2015), reducing yields and potentially leading to increases in food prices (World Bank 2012). Droughts and heat waves result from climate variability, but climate change may increase their frequency, severity, and other characteristics (IPCC 2013).

Multiple studies have explored the potential future changes in extreme events (Orlowsky and Seneviratne 2013; Sillmann et al. 2013; Maloney et al. 2014; Wuebbles et al. 2014), including droughts (Sheffield and Wood 2008; Dai 2011; Trenberth et al. 2013; Jeong et al 2014; Cook et al. 2015; Touma et al. 2015) and heat waves (Abatzoglou and Barbero 2014; Russo et al. 2014) over North America. These were based on climate model experiments from the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al. 2012) that informed the Intergovernmental Panel on Climate Change 5th Assessment Report (IPCC 2013). Past work has also looked at impacts of droughts and heat waves on agriculture (Mishra and Cherkauer 2010; Lobell et al. 2013; Lobell et al. 2014) and how this sector may be affected over North America under different climate change scenarios (Parry et al. 2004). While common trends have been identified, such as the drying of the

US Southwest and rising air temperatures throughout North America, there is still high uncertainty in the future projections (Allen et al. 2000; Knutti et al. 2008; Knutti and Sedláček 2013), especially regarding extreme events (Burke and Brown 2008; Sheffield and Wood 2008).

This uncertainty results from a variety of sources, including internal variability of the climate system (Deser et al. 2014), the degree of future mitigation of anthropogenic greenhouse gases (Diffenbaugh and Giorgi 2012), and the climate models used (Knutti et al. 2010; Cheruy et al. 2014; Friedlingstein et al. 2014). The relative contribution of each of these uncertainty sources to the overall value depends on the time horizon of the projections. For example, internal variability dominates uncertainty in the present and can have important contributions even up to 50 years into the future (Thompson et al. 2015), while uncertainties regarding emissions and climate models play an increasing role through the end of the century (Hawkins and Sutton 2009). Internal variability is difficult to predict because it arises from complex interactions within the climate system. Similarly, it is challenging to predict greenhouse gas emissions because they depend on human development and mitigation efforts. Uncertainty from climate models occurs because they include different sets of physical processes, use different parameterizations, or have different spatial and vertical resolutions, even though they share significant components and seek to solve the same general physical equations (Knutti et al. 2013).

The work presented here focuses on the uncertainty in future projections of droughts and heat waves derived from the diversity of climate models in CMIP5, and seeks to constrain it by using information on the models' historical biases. These biases are calculated using observationally constrained model output for key variables of the land surface, near-surface atmosphere, and their interactions, which are important in representing and controlling the occurrence of droughts, heat waves, and their feedbacks. Land-atmospheric (L-A) coupling is one of the main physical processes examined. This represents how much influence the land surface has on the lower part of the atmosphere and vice versa. The type of coupling determines whether a region is water-limited (evapotranspiration is positively correlated with soil moisture), energy-limited (evapotranspiration is negatively correlated with soil moisture), or in a transition zone (Seneviratne et al. 2010), with important implications for the occurrence of droughts and heat waves. L-A coupling can intensify droughts and increase their persistence (Wu and Kinter 2009; Roundy et al. 2013; Roundy et al., 2014), and generate and strengthen local heat waves (Fischer et al. 2007a,b; Lorenz et al. 2010; Berg et al. 2014; Miralles et al. 2014). Compound events, where droughts and heat waves take place simultaneously, cause large damages to crops due to both water and heat stress (Lesk et al. 2016). It is expected that L-A coupling will become more important in the future under climate change, especially for regions under transitional and dry regimes (Dirmeyer et al. 2013a,b). If this is the case, a stronger feedback between the land surface and the atmosphere may lead to increased drought persistence and intensity, and frequency of compound events.

An accurate depiction of the historical climate is a necessary, albeit not sufficient, condition to have confidence in the projections of a given climate model (Tebaldi and

Knutti 2007). For example, models that have positive temperature biases in the historical period have been shown to project larger increases in temperatures (Cheruy et al. 2014) because they generally overestimate incoming shortwave radiation due to misrepresentation of cloudiness. Furthermore, biases in L-A coupling strength can have an important impact on models' future projections. If a model displays stronger coupling, more incoming radiation will heat the lower atmosphere, especially during dry soil moisture periods (Seneviratne et al. 2010; Jaeger and Seneviratne 2011). This may then lead future increases in net radiation from increased CO_2 to be exaggerated. Conversely, if a model is too wet and coupling too weak, potential trends in desertification, droughts, and heat waves will be underestimated due to dampening of these land-atmospheric feedbacks.

The historical biases are used to develop model rankings based on the climate models' historical performance. Many studies have analyzed historical biases in climate models (e.g. Reichler and Kim 2008; McCrary and Randall 2010; Sheffield et al. 2013a,b) but have generally not linked performance to uncertainty in future projections. Several studies have also sought to develop model rankings to inform ensemble means where the contribution of each model depends on its performance (Brekke et al. 2008; Gleckler et al. 2008; Santer et al. 2009) instead of the more common "one model, one vote" criterion. However, not much emphasis has been placed on constraining the uncertainty of droughts (Wehner et al. 2011) and heat waves in particular, nor do past studies examine how the choice of performance metrics affects the resulting uncertainty of future projections.

2. DATA AND METHODS

2.1. Data

Data from 24 CMIP5 models from 14 modeling centers were used and are listed in Table 1 along with their general characteristics. The models were chosen based on the availability of variables needed for this study, in particular soil moisture at different layers, or with a total soil column under 2.5 meters. The historical (~1850-2005) experiment simulations were used to evaluate the models' biases, and the Representative Concentration Pathway 8.5 (RCP 8.5; Vuuren et al. 2011) simulations (~2006-2100) to explore the biases' relationships with future projections.

Observational data and observation driven land surface hydrological model output were taken from the North American Land Data Assimilation System Phase 2 (NLDAS-2; Xia et al. 2012a,b). The NLDAS-2 runs multiple land surface models over the continental US at 1/8th degree spatial and 1-hour temporal resolution for 1979 to present, in support of understanding the land surface hydrological cycle, drought monitoring and forecasting, and initialization of weather models (Xia et al. 2012a). The NLDAS-2 data have been evaluated against a range of observations, including streamflow (Xia et al. 2012b), soil moisture (Xia et al. 2014), soil temperature (Xia et al. 2013), and evapotranspiration (Peters-Lidard et al. 2011). It provides arguably the best estimate of land-surface hydrology at high resolution for the contiguous US, in particular for soil moisture and evapotranspiration, for which direct observations are lacking over large-scales and long time periods (> 10 years) (Nearing et al. 2016). Data for two of the NLDAS-2 models were used to evaluate the CMIP5 historical run climatologies: the Variable Infiltration Capacity model (VIC; Liang et al. 1994) and the Noah model (Chen et al. 1996). These two models were chosen because they provided the best overall performance in the evaluation studies mentioned previously. Data from these two models were averaged to produce the NLDAS-2 estimates.

The common time period of 1979-2005 was chosen for the comparisons between the NLDAS-2 and the CMIP5 historical data. The future changes were calculated between the end of the 21st century, 2070-2099 and this historical time period. Data from CMIP5 and NLDAS-2 models were interpolated to the grid with the lowest resolution amongst the models (i.e. 2.8x2.8 degrees).

2.2. Definition of droughts and heat waves

There are multiple definitions of a drought (Wanders et al. 2010; Sheffield and Wood 2011; Lloyd-Hughes 2013), and the decision of which to use depends on the application. We focus on summer agricultural drought calculated from monthly soil moisture (SM, kg/m^2/month) for June, July, and August (JJA), for a standard depth of 2 meters. Some of the models only report soil moisture for a total soil column depth between 1.5 – 2.5 meters, which was used directly. Other models reported data for multiple soil layers, so these were interpolated to a 2-meter level, assuming that soil moisture varied linearly between layers.

We carried out tests (not shown) using data from models that reported soil moisture at multiple layers to understand the impact of using values for slightly shallower (e.g. 1.5 meters) or slightly deeper (e.g. 2.5 meters) columns. Projected changes in drought frequency, duration, and severity (defined in Table 2) were calculated over the Crop Area (defined in Figure 2) for twelve models at all their reported depths. All models projected increases in drought frequency overall. However, three models projected higher increases in the probability of a drought occurring as a function of depth at an average rate of 5% per meter. On the other hand, nine models showed decreased changes in drought frequency as a function of depth with an average rate of -8% per meter. Nine models projected more severe events as a function of depth, driven mainly by increased drought duration in deeper soil columns. This suggests that models with deeper soil columns will tend to underestimate changes in drought frequency and overestimate their severity compared to a 2-meter baseline. How much so depends greatly on the model.

To quantify and understand the historic biases and future changes in soil moisture (and hence drought), three other variables were considered: monthly precipitation (Prcp, kg/m^2/month), evapotranspiration (ET, kg/m^2/month), and near-surface air temperature (Tas, K). JJA climatologies for the historic and future periods were calculated for all variables. The winter (December, January, February or DJF) and spring (March, April, May or MAM) climatologies were also calculated for precipitation because summer droughts are related to the previous seasons via snowpack and soil moisture persistence. Daily maximum near-surface air temperature (Tasmax, K) was used to identify heat waves (Lau and Nath 2012). Since heat waves usually last on the order of days to weeks, daily data between June 1st and August 31st were used.

Drought events were calculated from monthly soil moisture fields over a depth between 1.5 and 2.5 meters, depending on the model. An empirical cumulative distribution function (ECDF) was calculated for each summer month (i.e. June, July, August) for the historical period for each grid cell, and was used to calculate a percentile value for each month throughout the record. A month was defined to be under drought if soil moisture was below the 20th percentile (Sheffield et al. 2009). For future projections, the ECDF of the historical period was used to calculate the equivalent percentile for the future soil moisture values, thus including any shifts in the climatology as well as changes in variability.

There are also several definitions of heat waves (Robinson 2001; Della-Marta et al. 2007; Fischer et al. 2007a; Anderson and Bell 2011; Lau and Nath 2012). It is common to use a fixed value threshold for a given number of consecutive days (e.g. above 30 ºC for five days) (Della-Marta et al. 2007). This has the advantage of being easily translated to agricultural impacts where these thresholds have been linked to reduced yields (e.g. Lobell et al. 2013). Nevertheless, this type of definition poses a challenge when using CMIP5 data because models have temperature biases, leading to under or overestimation of heat waves relative to observations depending on the sign of the biases. Another definition of heat waves is based on percentiles (e.g. Anderson and Bell 2011), similar to our definition of soil moisture drought. This has the advantage of bypassing biases in temperature by defining the extreme events relative to the climatology of each model. For this reason, the latter method was chosen. Heat waves were calculated based on Tasmax, when values were above the 80th percentile (for consistency with the drought analysis) for five consecutive days, based on the ECDF for each day in JJA. For future heat waves, the historical ECDF was used to calculate the percentile values.

Yearly frequency, mean duration, mean intensity, and mean severity were calculated for both droughts and heat waves. The respective equations are defined in Table 2.

2.3. Definition of land-atmosphere coupling strength

Land-atmosphere processes depend largely on the type and strength of the dependence of evapotranspiration on soil moisture (Seneviratne et al. 2010), and whether it is water-limited, radiation-limited, or transitional. The strength of the coupling is also modulated by the magnitude of evapotranspiration. For example, in dry regions, the correlation between soil moisture and evapotranspiration is large and positive. However, as evapotranspiration is generally low, there is little feedback with the atmosphere. Therefore, strong L-A interactions take place where there is a combination of strong positive correlation between soil moisture and evapotranspiration, and a relatively high evapotranspiration rate.

Several metrics have been proposed to quantify the type and strength of L-A coupling and how they are represented in climate models (Koster et al. 2002; Dirmeyer 2006;

Dirmeyer et al. 2006). One metric commonly used in the literature (Dirmeyer et al. 2013b) is the correlation of interannual evapotranspiration (or latent heat flux) and soil moisture ρ(SM,ET), multiplied by the interannual standard deviation of evapotranspiration σ(ET), shown in Equation 1:

$$\gamma = \sigma(\mathrm{ET})\rho(\mathrm{SM,ET}) \qquad (1)$$

The correlation identifies if a region is typically water or energy limited over a timespan of decades. The standard deviation multiplier adds information about the variability of the evaporative flux throughout the data record. Thus, the metric quantifies the variability of land-atmospheric coupling strength within a region from year to year. However, for this study, it is more important to use the evapotranspiration climatology (the average of the flux's strength) instead, to capture the regions where, generally, the land surface has the capability of impacting the atmosphere within the summer season.

Figure 1 plots the mean JJA evapotranspiration against the interannual standard deviation of JJA evapotranspiration for water-limited, radiation-limited, and transition regions in the domain of the two NLDAS-2 models. Grid-cells are defined as water-limited when the correlation between soil moisture and evapotranspiration is significant ($p<0.05$) and larger than 0.3, as radiation-limited when this correlation is significant and negative with a magnitude larger than 0.3, and as transition otherwise. While this threshold is arbitrary, the sensitivity of the grid-cell classification to it is low until high thresholds (R=0.5-0.8) are chosen. This shows that there are water-limited regions with high mean values and low standard deviations, as well as regions with low mean values and high standard deviations. Therefore, to better account for seasonal L-A feedbacks, we modify the coupling metric by replacing the interannual standard deviation by the mean value μ(ET), and normalizing by the maximum evapotranspiration value throughout the domain max(ET), as shown in Equation 2:

$$\phi = \frac{\mu(\mathrm{ET})}{\max(\mathrm{ET})}\rho(\mathrm{SM,ET}) \qquad (2)$$

This normalization bounds the metric between -1 (strongly radiation-limited) and 1 (strongly water-limited).

The spatial patterns of interannual correlations between JJA soil moisture and evapotranspiration are very similar for both Noah and VIC (not shown), with the largest difference over the Southeast, where Noah shows higher mean evaporative fluxes compared to VIC. To account for the uncertainty of these estimates, we averaged the model values to a single NLDAS-2 ensemble mean. The percentage errors of the difference between the models' estimates with respect to the ensemble mean were calculated for climatologies in JJA SM, JJA ET and JJA ρ(SM,ET) (both models have the same meteorological forcings). These were found to be 41%, 33%, and 43%, respectively when averaged over the entire NLDAS-2 domain. Figure 2 shows the NLDAS-2 ensemble mean of the two coupling metrics given by Equations 1 and 2, i.e. γ and ϕ, respectively. Note that γ was normalized by the maximum standard deviation value in the domain to allow for comparison between the two. Here, γ shows a lower coupling in the US Southeast (a relatively wet region) than over the North of Mexico (a semi-arid region), in contrast to ϕ. Given that land-atmospheric coupling depends heavily on the strength of evaporative fluxes, which in turn depend on water availability, one would expect higher coupling over the US Southeast compared to the North of Mexico. For the rest of the study, the domain is split into seven sub-regions based on the spatial patterns of L-A coupling shown by ϕ.

L-A coupling is also a function of soil moisture depth, given that evapotranspiration takes place in the upper region of the soil column, depending on the distribution of the vegetation's roots (Rodríguez-Iturbe and Porporato 2004). The 1.5-2.5 meters depth generally encompasses the root zone and is deep enough to capture longer-lasting soil moisture memory beyond the frequency of individual storm events. This is in contrast to the upper soil layer (e.g. 10 cm), which experiences fluctuations at a higher frequency and therefore does not represent L-A coupling accurately at monthly time scales. However, soil columns between 2 and 2.5 meters can be deep enough to dampen some of the coupling strength if the vegetation has shallower roots in a given region. As with the droughts statistics, we explored the sensitivity of ϕ to soil depth, and found different sensitivities across models. Nine models showed an expected decrease in coupling with an average change of 16% per meter relative to the 2-meter value, whilst three models surprisingly showed an average increase in coupling with soil depth of 2% per meter.

2.4. Definition of sub-regions

We define a set of sub-regions that captures the spatial variation in L-A coupling. Figure 2 shows the coupling metric calculated from the average of the NLDAS-2 models. The Southeast shows the strongest coupling strength in the domain. The Northeast has negative coupling values as the region is wet and strongly radiation-limited. The

Northwest and Southwest have strong positive correlations between soil moisture and evapotranspiration since they are generally drier regions. However, L-A coupling is low since the seasonal evapotranspiration is also low. An important agricultural region "Crop Area" (Bagley et al. 2012) is further split into "Crop Upper" and "Crop Lower" because the difference in their coupling may have different implications for future changes.

2.5. Estimation of historical biases and relationship with future projections

Historical biases in each variable are calculated relative to the NLDAS-2 data, by averaging the data over each sub-region and subtracting the NLDAS-2 estimates from the CMIP5 model estimates. In this study the focus is on relating the biases in JJA Prcp to future projected changes via linear regression across models for each sub-region. This assumes that there is a linear relationship between the projected changes and the predictor. However, the biases in different variables are not independent: for example, biases in Prcp are associated with biases in ET in water-limited regions. We quantify this dependency by calculating the correlation matrices between the biases for each region across models.

3. RESULTS

3.1. Historical biases in mean climate

Figure 3 (a)-(f) show boxplots of the historical biases of MAM and JJA Prcp, JJA ET, JJA Tas, JJA ρ(SM,ET), and JJA land-atmospheric coupling metric ϕ across the 24 climate models averaged over each sub-region. Additionally, Table 3 lists the biases for each model for DJF, MAM and JJA Prcp, JJA ET, Tas, and ϕ for the Crop sub-regions. Of all regional biases, 61.3% were statistically significant ($p<0.05$) using a two-sample T-test. Over the Northeast, Northwest, Southwest, and Southeast the CMIP5 models show median positive biases for MAM Prcp amounting to a median percentage error of 30%, 33%, 76%, and 13%, respectively. Median biases were also found to be positive for JJA Prcp in these regions, with respective median percentage errors of 18%, 49%, 24%, and 8%, respectively. All four regions show median positive biases in ET (median percentage errors of 35%, 37%, 47%, and 18%). These four sub-regions also have small negative biases in Tas (median percentage errors of 0.44%, 0.84%, 0.55%, and 0.51%, respectively). The Northeast shows a median positive bias in ϕ (median percentage errors of 532%), while the Northwest, Southwest, and Southeast show a median negative bias (median percentage errors of 24%, 32% and 29%). In the Northeast, both components of ϕ are generally overestimated such that 21 of the models do not represent this region as being radiation-limited, resulting in such a large percentage error. In the Northwest, Southwest and Southeast, ϕ is underestimated by the median of the CMIP5 models because ρ(SM,ET) is underestimated. Here there are probably two competing effects: models with positive biases in Prcp represent these regions as being less water-limited, decreasing ρ(SM,ET), while their positive biases in ET increase μ(ET). Biases in ϕ are then a result of these effects on each of its components over each sub-region. Except for the case of Tas for which the climate models represent the climatologies quite well, the hydrological variables are relatively poorly represented by the median of the models.

These biases are not independent from each other. The cross-correlations between biases in each variable across the 24 models are shown in Figure 3(g)-(n). Biases in DJF Prcp are not shown but were positively correlated with those in MAM Prcp everywhere except for the Southeast, and with those in JJA Prcp and JJA ET in the Northwest. They were also negatively correlated with biases in Tas over the Southwest. Models that have higher JJA Prcp also tend to have higher MAM Prcp (except in the Southeast and Crop areas), higher JJA ET, lower JJA Tas (except in the Northwest and Southwest), and lower correlations between SM and ET (except in the Southwest). The lower temperatures are consistent with a wet bias that induces more ET and more evaporative cooling. Conversely, models with less summer Prcp also tend to experience a drier spring, lower ET, higher Tas, and a stronger dependence of ET on SM. Interestingly, no region showed a significant correlation ($p<0.05$) between biases in JJA Prcp and biases in ϕ. This is probably because of the competing effects mentioned in the previous paragraph, whereby higher Prcp leads to higher ET rates but also lower ρ(SM,ET), thus having mixed effects on ϕ. Low correlations in other regions may also be related to how the models represent ET and SM dynamics, irrespective of the biases in Prcp. Overall, the correlations show that there are common climate regimes for the historical period across the models: models that are wetter (drier) during the summer, are also wetter (drier) in the spring, have higher (lower) ET, lower (higher) Tas, and weaker (stronger) relationships between ET and SM.

3.2. Relationship between historical biases and future projected changes in mean climate

The ranges of projected changes in MAM and JJA Prcp, JJA ET, Tas, ρ(SM,ET), and ϕ from the 24 climate models are shown in Figure 4 (a)-(f). Furthermore, Table 4 lists the projected changes for each model for DJF, MAM, and JJA Prcp, JJA ET, Tas, and ϕ over the Crop sub-regions. The ranges are large and there is no absolute consensus on the sign of most of these changes across regions. The median of the models show an increase in MAM Prcp in every sub-region but the Southwest, while the median also shows slight decreases of JJA Prcp, albeit with several models showing no changes or a positive one. These two changes are positively correlated across models (Figure 4(g)-(n)) because those that project the largest decreases in JJA Prcp also project decreases in MAM Prcp, and those that project no or positive changes in JJA Prcp, project increases in MAM Prcp. Changes in ET are more uncertain in the Southeast and the Crop Area, though most models project increases in the Northeast and Northwest, and decrease in the Southwest. All models and sub-regions show an increase in Tas with a median of 5.0 °C across the domain. However, some models project an increase of up to 8.5 °C over the Crop Upper region. This large disparity in projected changes in temperature has been partially attributed to the models' historical biases in incoming shortwave radiation due to misrepresentation of clouds. Models with the highest deficiencies in depicting cloudiness tend to project the largest temperature increases in midlatitude areas globally (Cheruy et al. 2014). The median of the models shows projected increases in the correlation between

SM and ET, and the coupling metric except for the Southwest, although there is large disagreement on the signs of these changes.

To understand how the projected changes in each variable are related, their cross-correlations were calculated across models for each sub-region (Figure 4(g)-(n)). Changes in DJF Prcp are not included as they were only positively correlated with changes in MAM Prcp over the Southwest, Southeast, and the Crop areas. There are several strong correlations for changes in temperature, which are negatively correlated with changes in JJA Prcp in the Northeast, Southwest, Southeast, and the Crop areas. This shows that by the end of the century, the models tend to fall into a range of climates over certain regions. On one hand, models with higher increases in JJA Prcp are likely to also have a wetter spring over the Southwest, Southeast, and Crop areas, higher JJA ET rates across regions, stronger ϕ (except in the Northeast and Northwest) and dampening the JJA Tas increase (except in the Northwest). Conversely, models that exhibit the highest increases in temperature also tend to experience the largest decreases in Prcp and ET, and a weakening of ϕ.

A linear regression was fitted between the historical biases in JJA Prcp and the projected changes in MAM and JJA Prcp, JJA ET, Tas, ρ(SM,ET), and ϕ across climate models and for each sub-region. Figure 5 displays the regression slopes and R^2 values (left panels), and intercepts (right panels). No significant relationships ($p<0.05$) were found with changes in DJF Prcp, so they are not shown.

Figure 5 shows that for the Northeast, Northwest, and the Crop Area, a positive bias in JJA Prcp is related to larger positive increases in MAM Prcp, amounting to 20%, 40%, and 18% of the variance in the model projections in each region, respectively. The same relationship is evident for changes in JJA Prcp over the Southeast and the Crop Area, though models with smaller bias (close to the regression intercept) project a decrease in Prcp in the Southeast (shown by the negative regression intercept) and no change over the Crop Area. The percentages of the variance explained by this relationship are 19% and 22%, respectively. For example, the regression slope and intercept of the projected changes in JJA Prcp against bias in JJA Prcp over the Crop Area are 0.26 mm month^{-1}/mm month^{-1} and -1.7 mm month^{-1}, respectively (p=0.036). This positive relationship between historical bias in JJA Prcp and its projected changes means that a wetter model during the historical period will tend to project a wetter US by the end of the 21st century if the bias is large, or little change in JJA Prcp if the bias is small.

Projected future changes in ϕ in the Northwest, Southeast, and the Crop Area also show significant positive relationships with biases in JJA Prcp, with percentage variances explained of 23%, 36%, and 28%, respectively. These are related to greater increases in ET rates (slope = 0.24 mm month^{-1}/mm month^{-1} increase over the Crop Area) and greater strengthening of ρ(SM,ET) (slope = 0.002 1/mm month^{-1} increase in the Crop Area) in historically wetter models. This last relationship is particularly interesting since wetter models during the historical period were found to be associated with weaker ρ(SM,ET) (Figure 3(g)-(n)). The relationships from Figures 4(g)-(n) show that these same wetter models project increases in ρ(SM,ET), albeit with a very shallow slope. Thus, wetter models during the historical period project a strengthening of the coupling due to increases in both components of ϕ in the future within the Southeast and the Crop Area, though especially due to that in ET. A possible explanation for this is that higher future temperatures will drive increases in ET such that the regions become more water-limited despite increases in Prcp. In turn, drier models during the historical period project a weakening of ϕ likely due to the decreases in ET associated with decreases in JJA Prcp, since the small slope of the correlation component suggests that it has little impact on the overall changes of ϕ. In the Northwest, models with little bias in JJA Prcp tend to project a decrease in JJA ET. A drier model in this region would then tend to project an even larger decrease and a wetter model a very small decrease, or even an increase in JJA ET if the JJA Prcp bias was large.

3.3. Implications for Extreme Events: Droughts and Heat Waves

Figures 6 and 7 show the changes in yearly frequency plotted against changes in mean severity of drought and heat wave events, respectively. There are strong positive relationships between changes in drought frequency and severity throughout every sub-region, with R-values ranging from 0.64 in the Northwest and Crop Area, to 0.82 in the Northeast. Therefore, models that show the highest increases in the number of droughts relative to the historical period also experience larger increases in drought severity, which is to be expected given the use of a fixed percentile based threshold.

For example, MIROC5, MIROC-ESM, and MIROC-ESM-CHEM project the largest increases in drought frequency over the Crop Area, together with soil moisture drying (not shown). This is likely driven by their projected reductions in JJA rainfall (-10.6, 12.6, and 10.2 mm/month, respectively) over this area relative to changes in ET (-0.2, -1.6, -0.6 mm/month, respectively), as shown in Table 4. Additionally, Table 3 displays that two of them have large negative biases in JJA Prcp (-3.9, -18.8, -15.1 mm/month, respectively).

Figure 7 shows that models exhibit a positive relationship between increases in heat wave frequency and severity throughout the domain, with the strongest correlation (R=0.58) over the Northeast. Two models that project large increases in heat wave frequency and severity are MIROC-ESM and GFDL-CM3. Both these models project higher changes in daily maximum and monthly values of near-surface air temperature (not shown). Given the projected changes and biases in MIROC-ESM already discussed, its projected increases in heat waves are possibly due to the increased partitioning of incoming radiation into sensible heat flux. GFDL-CM3, on the other hand, has a small positive bias in JJA Prcp of 5.9 mm/month and projects an increase in Prcp (8.3 mm/month) and ET (21.5 mm/month). In this case, it could be that larger-scale factors are responsible for the higher increases in temperature (7.2 K compared to the 24 model ensemble increase of 5.3 K). Another possible explanation is that there might be changes in the distribution of rainfall throughout the summer, which might leave longer drier periods that might encourage the formation of heat waves.

A Spearman rank correlation was calculated between the absolute projected changes in each of the characteristics of droughts and the historical biases in JJA Prcp. This was repeated for the changes in heat waves and biases in JJA Tas. The results are shown in Figure 8. Drought yearly frequency has significant ($p<0.05$) negative correlations with biases in JJA Prcp over the Northeast, Southeast, the Crop Area and Crop Upper. Drought mean intensity has similar negative correlations with biases in JJA Prcp over the Northeast and the Southeast. Drought mean duration is also correlated with biases in JJA Prcp over the Northeast, Southeast, Crop Area, and Crop Lower. Finally, drought mean severity is negatively correlated with biases in JJA Prcp over every region except for the Northwest and Southwest. These results show that wetter models during the historical period tend to project less frequent, less intense, and shorter droughts, while drier models will produce more extreme projections of these drought characteristics in many of the sub-regions, particularly over those important for agriculture.

Fewer significant Spearman rank correlations were found between biases in JJA Tas and changes in heat wave characteristics (and none with biases in JJA Prcp). These biases are correlated with changes in heat wave yearly frequency over the Northeast and Crop Upper. Changes in mean intensity are also correlated with these biases over Crop Upper and the Northwest. Significant relationships were found for the changes in heat wave mean duration and mean severity, but solely over the Northwest. While these relationships are fewer, given the correlation between biases in JJA Tas and JJA Prcp we can infer (albeit rather weakly) that drier and hotter models produce more extreme projections for heat waves, mainly over the Northwest region, compared to those models that tend to be wetter and cooler over the historical period.

4. DISCUSSION AND CONCLUSIONS

4.1. Potential constraints on the uncertainty of future projections

The question remains whether it is possible to use the information on the model biases to constrain the uncertainty of future projections. In this section we use the biases to rank the models, assuming that an accurate representation of the historical climate is necessary (albeit not sufficient) for trusting the projected changes in future hydroclimate and its extremes. There is an incentive to develop these model rankings because climate change impact studies often select a small subset of the climate models on which to base their analyses (e.g. Brekke et al. 2009; Schewe et al. 2014). Since small subsets of climate models are driving the community's research on the potential impacts of climate change (e.g. Gerten et al., 2011; Hagemann et al., 2011; Warszawski et al., 2014; Frieler et al., 2015), one would desire for the "best" models to be used, whilst encompassing a realistic range of uncertainty for the timeframe of interest (e.g. near-term, mid-century, end-of-century). Constraining the uncertainty that arises from model diversity is important because it represents the largest contribution of the overall uncertainty of climate change by the end of the century for a given RCP scenario (Hawkins and Sutton 2009).

The 24 models were ranked according to the absolute values of their biases in JJA Prcp, JJA Tas, JJA ET, and JJA ϕ. These rankings were done separately for each variable. A Spearman correlation analysis between the rankings showed positive significant correlations ($p<0.05$) between those from JJA Prcp and JJA ET ($R=0.53$), and those from JJA Tas and JJA ET ($R=0.45$). A negative and significant correlation was found between the rankings derived from JJA ET and JJA ϕ ($R=-0.57$). More details can be obtained from Table 3. The lack of more correlated rankings is possibly because negative biases are treated the same as positive ones and because a discrete ranking may amplify the differences between models with statistically similar biases.

To understand the error in the uncertainty range that derives from selecting a subset of the 24 models, we randomly sampled subsets of models and compared their ranges of projections to those when selecting the top performing models according to the rankings. This was done using bootstrap sampling whereby a subset of models was selected at random 1,000 different times from the ensemble of 24 climate models. The interquartile range of the projected changes in droughts and heat waves was calculated for each sample as a measure of uncertainty. This sampling was done for sample sizes from 5 to 23 models to quantify how this uncertainty range changes as a function of the sample size. In parallel, subsets of models (from 5 to 23 models) were selected according to the four rankings over the Crop Area and the interquartile range of their projected changes calculated. This allowed us to compare the uncertainty derived when selecting the "better performing" models as opposed to selecting the same number of models at random.

The results of this uncertainty analysis are displayed in Figure 9. The median of the bootstrap analysis shows that selecting a small sample of models (e.g. 5) at random will likely underestimate the variance of the projected changes compared to that from the 24 models. Selecting a small sample of models using the rankings based on JJA Prcp, Tas, and ϕ yields overall larger uncertainty ranges for the projected changes in drought yearly frequency than the median of the bootstrap analysis. Conversely, the ranking from JJA ET consistently produces a lower uncertainty range. For the changes in drought severity, all the rankings lead to lower uncertainty ranges for most of the model samples, although the rankings from JJA Prcp and ET approach the median value from the bootstrap for samples larger than 13 models. For the changes in heat wave frequency, all the rankings consistently yield higher uncertainty ranges than the bootstrap median. However, they all lie close to the bootstrap median when analyzing the projected changes in heat wave severity.

This analysis suggests that selecting small subsets of the CMIP5 models will most likely artificially reduce the uncertainty range of the projections in question (Knutti et al. 2010) regardless of how the models are chosen. It also reiterates the challenge of developing consistent model rankings (e.g. Gleckler et al. 2008), even with a particular application in mind (in this case, to study droughts and heat waves). Here we show that even when historical biases in hydroclimatic variables account for some of the variability of projections across models, it is not enough to generate consistent model rankings that can constrain the projections' uncertainty ranges.

Without being able to determine the “best” models in a logical and rigorous way, it might be more appropriate to span the full range of model uncertainty, as we know it. This would allow for a more accurate characterization of the potential impacts of climate change. As shown by Figure 9, it is possible in some cases to increase the likelihood of matching the full uncertainty range using a large enough subset (e.g. 10 models). However, the uncertainty may still be under- or overestimated depending on the subset. Studies that use a small number of climate models chosen arbitrarily should be cautious in their conclusions, since they are likely underestimating the range of possible outcomes resulting from climate change by artificially selecting a small subset. Model uncertainty is an important component of the overall uncertainty estimates of climate change both at short and long time scales, so it should not be neglected by arbitrarily choosing a small number of models.

4.2. Caveats

While there are more models available in the CMIP5 archive than the 24 that were analyzed here, they were not selected because they did not report soil moisture content at different layers, had a total soil column deeper than 2.5 meters, or were not readily available from the CMIP5 data portal. The selection of 24 models may underestimate the full uncertainty range from the CMIP5 models, as indicated by the sub-sampling experiments. Nevertheless this is likely to be small since there are decreasing marginal returns in added uncertainty as more models are added after around 10-15 models (Knutti et al. 2010; Ferro et al. 2012). A key question is whether the full CMIP5 ensemble of models represents the true uncertainty, or whether further diversity in the models is needed in terms of which processes are represented and how (Tebaldi and Knutti 2007; Knutti et al. 2010).

Linear relationships were found between historical biases and future projections, though the percentage of variance explained was relatively low for most variables and regions. This shows that using the climatologies of hydroclimatic variables to generate model rankings is not effective enough to reduce the uncertainty ranges, since there are many other factors involved. Moreover, the historical biases considered here were calculated from the limited time period of 1979-2005 that spans 27 years, so decadal variability is not fully captured by these climatologies leading to uncertainty in the calculated biases.

Nevertheless, the relationships of the historical biases on the models' future projections also show that simply removing the historical bias from future projections data will not be enough to remove the effects that a model's historical biases has on its resulting projections. More advanced statistical bias correction methodologies (e.g. Li et al. 2010; Hagemann et al. 2011) take into account the full distribution of the variables using quantile matching. However, future bias correction studies should also take into account the relationships between historical biases and projected changes that were explored here.

4.3. The role of land-atmospheric coupling

We show that there are significant biases across models in our chosen coupling metric that manifest in misrepresentation of whether a region is water-limited or radiation-limited as well as the magnitude of evapotranspiration. This study agrees with previous ones that have found that L-A coupling may intensify in the future over a large part of the US (Dirmeyer et al. 2013a,b). Depending on the main control of evapotranspiration in a region, the effect of strengthening L-A coupling would be different. For example, the projected increase in coupling strength in the Southeast and the Crop Area, which are already water-limited, could help drive the increase in drought persistence and severity. It could also lead to higher local increases in near-surface air temperature, leading to more frequent and intense heat waves and compound events. These potential increases in extreme events pose high dangers to future agriculture in the region.

4.4. Conclusions

This study quantified the biases of 24 CMIP5 models for precipitation, evapotranspiration, near-surface air temperature, and land-atmospheric coupling over the US. The ensemble of models tends to be biased wet and cool in most of the country and dry and warm in the Southeast for 1979-2005. These biases were linked to projected changes in the climatologies of hydrometeorological variables and extreme events under the RCP 8.5 scenario by the end of the 21st century. The wetter the models are during the historic period, the wetter they tend to project the end of the century to be due to larger increases in precipitation, and vice versa. This study finds stronger relationships between historical biases over the US, compared to the results of Knutti et al. (2010), carried out at a global scale. However, in most cases the relationships found in this work only accounted for a small fraction of the observed variance across models.

Most models agree on a general drying trend in soil moisture by the end of the 20^{th} century, and therefore more frequent and severe droughts are expected in the future. There is a wide range of projected changes that were often inversely correlated with historical biases in precipitation, such that wetter (drier) models projected smaller (larger) changes in drought characteristics. However, changes in DJF Prcp were significantly correlated with changes in droughts (not shown), but few relationships were found between this and other changes or with the historical biases, showing that there are other factors involved in the projected changes in droughts. All models show a positive shift in near-surface air temperature towards higher temperatures by the end of the century. Given these changes, all models project increases in heat wave frequency and severity, with large uncertainty across models. To a lesser degree, this range of projected changes in heat wave characteristics was also related to historical biases in near-surface air temperature.

This work has reiterated the challenge of constraining the uncertainty of future projections of droughts and heat waves. Here the focus was on the US, though it is likely that similar results would be obtained for other regions. There are, however, some changes with which most of the models in this study agree: there will be more frequent and severe droughts in the Southwest and the Southeast, and heat waves throughout the US by the end of the century if we follow the path given by the RCP8.5. The uncertainty lies mainly in the magnitude of these changes, rather than on their direction. Further attempts to constrain model uncertainty may focus instead on model performance at the process level, providing more insights on the origins of biases in climatologies used here. In the meantime, until a robust methodology to rank climate models is developed, researchers should aim to include more climate models in their impacts studies to characterize the possible range of projections more accurately.

6. ACKNOWLEDGEMENTS

We thank the members of the Terrestrial Hydrology Research Group at Princeton University, in particular Dr. Niko Wanders, and three anonymous reviewers for their comments, which helped improve this study. This work was supported by the NOAA Climate Program Office (NA11OAR4310097).

7. REFERENCES

Abatzoglou, J. T., and R. Barbero, 2014: Observed and projected changes in absolute temperature records across the contiguous United States. *Geophys. Res. Lett.,* **41**, 6501–6508, doi:10.1002/2014GL061441.

Allen, M. R., P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. L. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620, doi:10.1038/35036559.

Anderson, G. B., and M. L. Bell, 2011: Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities. *Environ. Health Perspect*., **119**, 210–218, doi:10.1289/ehp.1002313

Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.,* **38**, L05805, doi:10.1029/2010GL046270.

Bagley, J. E., A. R. Desai, P. A. Dirmeyer, and J. A. Foley, 2012: Effects of land cover change on moisture availability and potential crop yield in the world's breadbaskets. *Environ. Res. Lett.*, **7**, 14009, doi:10.1088/1748-9326/7/1/014009.

Bao, Q., and Coauthors, 2013: The Flexible Global Ocean-Atmosphere-Land System Model, spectral version 2: FGOALS- s2. *Adv. Atmos. Sci.,* **30**, 561–576, doi: 10.1007/s00376-012-2113-9

Berg, A., and Coauthors, 2014: Interannual coupling between summertime surface temperature and precipitation over land: processes and implications for climate change. *J. Climate*, 1308-1328, doi:10.1175/JCLI-D-14-00324.1.

Bi, D., and Coauthors, 2013: The ACCESS coupled model: description, control climate and evaluation, *Aust. Met. Oceanogr. J*., **63**, 41-64.

Brekke, L. D., M. D. Dettinger, E. P. Maurer, and M. Anderson, 2008: Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments. *Climatic Change,* **89**, 371–394, doi:10.1007/s10584-007-9388-3.

——, E. P. Maurer, J. D. Anderson, M. D. Dettinger, E. S. Townsley, A. Harrison, and T.

Pruitt, 2009: Assessing reservoir operations risk under climate change. *Water Resour. Res.*, **45**, W04411, doi:10.1029/2008WR006941.

Burke, E. J., and S. J. Brown, 2008: Evaluating Uncertainties in the Projection of Future Drought. *J. Hydrometeorology*, **9**, 292–299, doi:10.1175/2007JHM929.1.

Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268, doi:10.1029/95JD02165.

Cheruy, F., J. L. Dufresne, F. Hourdin, and A. Ducharne, 2014: Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 6493–6500, doi:10.1002/2014GL061145.

Cook, B. I., T. R. Ault, and J. E. Smerdon, 2015: Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci. Adv.*, **1**, 1-7, doi:10.1126/sciadv.1400082.

Dai, A., 2011: Drought under global warming: a review. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 45–65, doi:10.1002/wcc.81.

Dawadi, S., and S. Ahmad, 2012: Changing climatic conditions in the Colorado River Basin: Implications for water resources management. *J. Hydrol.*, **430**–**431**, 127–141, doi:10.1016/j.jhydrol.2012.02.010.

Della-Marta, P. M., J. Luterbacher, H. Weissenfluh, E. Xoplaki, M. Brunet, and H. Wanner, 2007: Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Climate Dyn.*, **29**, 251–275, doi:10.1007/s00382-007-0233-1.

Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American Climate over the Next 50 Years: Uncertainty due to Internal Variability. *J. Climate*, **27**, 2271–2296, doi:10.1175/JCLI-D-13-00451.1.

Diffenbaugh, N. S., and F. Giorgi, 2012: Climate change hotspots in the CMIP5 global climate model ensemble. *Climatic Change*, **114**, 813–822, doi:10.1007/s10584-012-0570-x.

Dirmeyer, P. A., 2006: The hydrologic feedback pathway for land-climate coupling. J. Hydrometeor., **7**, 857–867, doi: 10.1175/JHM526.1.

——, R. D. Koster, and Z. Guo, 2006: Do Global Models Properly Represent the Feedback between Land and Atmosphere? *J. Hydrometeor.*, **7**, 1177–1198, doi:10.1175/JHM532.1.

——, Y. Jin, B. Singh, and X. Yan, 2013a: Trends in Land–Atmosphere Interactions from CMIP5 Simulations. *J. Hydrometeor.*, **14**, 829–849, doi:10.1175/JHM-D-12-0107.1.

——, ——, ——, and ——, 2013b: Evolving Land–Atmosphere Interactions over North America from CMIP5 Simulations. *J. Climate*, **26**, 7313–7327, doi:10.1175/JCLI-D-12-00454.1.

Donner, L. J., and Coauthors, 2011: The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Climate*, **24**, 3484–3519, doi: 10.1175/2011JCLI3955.1

Dufresne, J.-L., and Coauthors, 2013: Climate change projections using the IPSL-CM5 earth system model: From CMIP3 to CMIP5. *Climate Dyn.*, **40**, 2123–2165, doi:

10.1007/s00382-012-1636-1

Ferro, C. A. T., T. E. Jupp, F. H. Lambert, C. Huntingford, and P. M. Cox, 2012: Model complexity versus ensemble size: allocating resources for climate prediction. *Phil. Trans. R. Soc. A*, **370**, 1087–1099, doi:10.1098/rsta.2011.0307.

Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär, 2007a: Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophys. Res. Lett.*, **34**, L06707, doi:10.1029/2006GL029068.

——, ——, P. L. Vidale, D. Lüthi, and C. Schär, 2007b: Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave. *J. Climate*, **20**, 5081–5099, doi:10.1175/JCLI4288.1.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti, 2014: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *J. Climate*, **27**, 511–526, doi:10.1175/JCLI-D-12-00579.1.

Frieler, K., and Coauthors, 2015: A framework for the cross-sectoral integration of multi-model impact projections: land use decisions under climate impacts uncertainties. *Earth Syst. Dynam.*, **6**, 447–460, doi:10.5194/esd-6-447-2015.

Gerten, D., J. Heinke, H. Hoff, H. Biemans, M. Fader, and K. Waha, 2011: Global Water Availability and Requirements for Future Food Production. *J. Hydrometeor.*, **12**, 885–899, doi:10.1175/2011JHM1328.1.

Giorgetta, M. and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572-597, doi:10.1002/jame.20038.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi:10.1029/2007JD008972.

Hagemann, S., C. Chen, J. O. Haerter, J. Heinke, D. Gerten, and C. Piani, 2011: Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *J. Hydrometeor.*, **12**, 556–578, doi:10.1175/2011JHM1336.1.

Hatfield, J. L., and J. H. Prueger, 2015: Temperature extremes: Effect on plant growth and development. *Wea. Climate Extremes*, **10,** 4–10, doi:10.1016/j.wace.2015.08.001.

Hawkins, E., and R. Sutton, 2009: The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, 1535 pp, doi:10.1017/CBO9781107415324.

Jaeger, E. B., and S. I. Seneviratne, 2011: Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model. *Climate Dyn.*, **36**, 1919–1939, doi:10.1007/s00382-010-0780-8.

Jeong, D. I., L. Sushama, and M. Naveed Khaliq, 2014: The role of temperature in drought projections over North America. *Climatic Change*, **127**, 289–303, doi:10.1007/s10584-014-1248-3.

Jones, C. D., and Coauthors, 2011: The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.*, **4**, 543–570, doi:10.5194/gmd-4-543-2011.

Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, **3**, 369–373, doi:10.1038/nclimate1716.

Knutti, R., and Coauthors, 2008: A Review of Uncertainties in Global Temperature Projections over the Twenty-First Century. *J. Climate*, **21**, 2651–2663, doi:10.1175/2007JCLI2119.1.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in Combining Projections from Multiple Climate Models. *J. Climate*, **23**, 2739–2758, doi:10.1175/2009JCLI3361.1.

——, D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, doi:10.1002/grl.50256.

Koster, R. D., P. A. Dirmeyer, A. N. Hahmann, R. Ijpelaar, L. Tyahla, P. Cox, and M. J. Suarez, 2002: Comparing the degree of land-atmosphere interaction in four atmospheric general circulation models. *J. Hydrometeor.*, **3**, 363–375, doi:10.1175/1525-7541(2002)003<0363:CTDOLA>2.0.CO;2.

Lau, N.-C., and M. J. Nath, 2012: A Model Study of Heat Waves over North America: Meteorological Aspects and Projections for the Twenty-First Century. *J. Climate*, **25**, 4761–4784, doi:10.1175/JCLI-D-11-00575.1.

Lesk, C., P. Rowhani, and N. Ramankutty, 2016: Influence of extreme weather disasters on global crop production. *Nature*, **529**, 84–87, doi:10.1038/nature16467.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, **99**, 14415–14428, doi:10.1029/94JD00483.

Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, doi:10.1029/2009JD012882.

Lloyd-Hughes, B., 2013: The impracticality of a universal drought definition. *Theor. Appl. Climatol.*, **117**, 607–611, doi:10.1007/s00704-013-1025-7.

Lobell, D. B., G. L. Hammer, G. McLean, C. Messina, M. J. Roberts, and W. Schlenker, 2013: The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, **3**, 497–501, doi:10.1038/nclimate1832.

Lobell, D. B., M. J. Roberts, W. Schlenker, N. Braun, B. B. Little, R. M. Rejesus, and G. L. Hammer, 2014: Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest. *Science*, **344**, 516–519, doi:10.1126/science.1251423.

Lorenz, R., E. B. Jaeger, and S. I. Seneviratne, 2010: Persistence of heat waves and its link to soil moisture memory. *Geophys. Res. Lett.*, **37**, L09703, doi:10.1029/2010GL042764.

Maloney, E. D., and Coauthors, 2014: North American Climate in CMIP5 Experiments: Part III: Assessment of Twenty-First-Century Projections. *J. Climate*, **27**, 2230–2270, doi:10.1175/JCLI-D-13-00273.1.

McCrary, R. R., and D. A. Randall, 2010: Great Plains Drought in Simulations of the

Twentieth Century. *J. Climate*, **23**, 2178–2196, doi:10.1175/2009JCLI3061.1.

Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. Vilà-Guerau de Arellano, 2014: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geosci.*, **7**, 345–349, doi:10.1038/ngeo2141.

Mishra, V., and K. A. Cherkauer, 2010: Retrospective droughts in the crop growing season: Implications to corn and soybean yield in the Midwestern United States. *Agric. and For. Meteor.*, **150**, 1030–1045, doi:10.1016/j.agrformet.2010.04.002.

Nearing, G. S., D. M. Mocko, C. D. Peters-Lidard, S. V. Kumar, and Y. Xia, 2016: Benchmarking NLDAS-2 Soil Moisture and Evapotranspiration to Separate Uncertainty Contributions. *J. Hydrometeor.*, **17**, 745–759, doi:10.1175/JHM-D-15-0063.1.

Orlowsky, B., and S. I. Seneviratne, 2013: Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections. *Hydrol. Earth Syst. Sci.*, **17**, 1765–1781, doi:10.5194/hess-17-1765-2013.

Parry, M. L., C. Rosenzweig, A. Iglesias, M. Livermore, and G. Fischer, 2004: Effects of climate change on global food production under SRES emissions and socio-economic scenarios. *Global Environ. Change*, **14**, 53–67, doi:10.1016/j.gloenvcha.2003.10.008.

Peters-Lidard, C. D., S. V. Kumar, D. M. Mocko, and Y. Tian, 2011: Estimating evapotranspiration with land data assimilation systems. *Hydrol. Process.*, **25**, 3979–3992, doi:10.1002/hyp.8387.

Reichler, T., and J. Kim, 2008: How Well Do Coupled Models Simulate Today's

Climate? *Bull. Amer. Meteor. Soc.*, **89**, 303–311, doi:10.1175/BAMS-89-3-303.

Robinson, P. J., 2001: On the definition of a heat wave. *J. Appl. Meteor.*, **40**, 762–775, doi:10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2.

Rodriguez-Iturbe, I. and A, Porporato, 2004: *Ecohydrology of Water Controlled Ecosystems: Soil Moisture and Plant Dynamics*. Cambridge University Press, 442 pp.

Roundy, J. K., C. R. Ferguson, and E. F. Wood, 2013: Temporal Variability of Land–Atmosphere Coupling and Its Implications for Drought over the Southeast United States. *J. Hydrometeor.*, **14**, 622–635, doi:10.1175/JHM-D-12-090.1.

——, ——, and ——, 2014: Impact of land-atmospheric coupling in CFSv2 on drought prediction. *Climate Dyn.*, **43**, 421–434, doi:10.1007/s00382-013-1982-7.

Russo, S., and Coauthors, 2014: Magnitude of extreme heat waves in present climate and their projection in a warming world. *J. Geophys. Res.: Atmos.*, **119**, 12500–12512, doi:10.1002/2014JD022098.

Santer, B. D., and Coauthors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Natl. Acad. Sci.,*, **106**, 14778–14783, doi: 10.1073/pnas.0901736106.

Schewe, J., and Coauthors, 2014: Multimodel assessment of water scarcity under climate change. *Proc. Natl. Acad. Sci.*, **111**, 3245–3250, doi:10.1073/pnas.1222460110.

Scoccimarro E., S. Gualdi, A. Bellucci, A. Sanna, P.G. Fogli, E. Manzini, M. Vichi, P. Oddo, and A. Navarra, 2011: Effects of Tropical Cyclones on Ocean Heat Transport in a High Resolution Coupled General Circulation Model. *J. Climate*, **24**, 4368-4384, doi: 10.1175/2011JCLI4104.1

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

Sheffield, J., and E. F. Wood, 2008: Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. *Climate Dyn.*, **31**, 79–105, doi:10.1007/s00382-007-0340-z.

Sheffield, J., K. M. Andreadis, E. F. Wood, and D. P. Lettenmaier, 2009: Global and Continental Drought in the Second Half of the Twentieth Century: Severity–Area–Duration Analysis and Temporal Variability of Large-Scale Events. *J. Climate*, **22**, 1962–1981, doi:10.1175/2008JCLI2722.1.

Sheffield, J., and E. F. Wood, 2011: *Drought: Past Problems and Future Scenarios*. Earthscan, 210 pp.

Sheffield, J., and Coauthors, 2013a: North American Climate in CMIP5 Experiments. Part I: Evaluation of Historical Simulations of Continental and Regional Climatology. *J. Climate*, **26**, 9209–9245, doi:10.1175/JCLI-D-12-00592.1.

——, and Coauthors, 2013b: North American Climate in CMIP5 Experiments. Part II: Evaluation of Historical Simulations of Intraseasonal to Decadal Variability. *J. Climate*, **26**, 9247–9290, doi:10.1175/JCLI-D-12-00593.1.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res.: Atmos.*, **118**, 2473–2493, doi:10.1002/jgrd.50188.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the

Experiment Design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc., A*, **365**, 2053–2075, doi:10.1098/rsta.2007.2076.

Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the Role of Internal Climate Variability in Future Climate Trends. *J. Climate*, **28**, 6443–6456, doi:10.1175/JCLI-D-14-00830.1.

Touma, D., M. Ashfaq, M. A. Nayak, S.-C. Kao, and N. S. Diffenbaugh, A multi-model and multi-index evaluation of drought characteristics in the 21st century. J. *Hydrol.*, **526**, 196-207, doi:10.1016/j.jhydrol.2014.12.011.

Trenberth, K. E., A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield, 2013: Global warming and changes in drought. *Nature Climate Change*, **4**, 17–22, doi:10.1038/nclimate2067.

Vliet, M. T. H. van, D. Wiberg, S. Leduc, and K. Riahi, 2016: Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nature Climate Change*, **6**, 375-380, doi:10.1038/nclimate2903.

Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dyn.,* **40**, 2091–2121, doi:10.1007/s00382-011-1259-y.

Volodin, E. M., N. A. Diansky, and A. V. Gusev, 2010: Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Izvestiya, Atmos. Oceanic Phys.,* **46**, 414–431, doi:

10.1134/S000143381004002X.

Vuuren, D. P. van, and Coauthors, 2011: The representative concentration pathways: an overview. *Climatic Change*, **109**, 5-31, doi:10.1007/s10584-011-0148-z.

Wanders, N., H. A. J. Van Lanen, and A. F. Van Loon, 2010: *Indicators for drought characterization on a global scale, WATCH Tech, Rep. 24*, Wageningen Univ., Netherlands, 54 pp. http://library.wur.nl/WebQuery/wurpubs/fulltext/160049.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe, 2014: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. *Proc. Natl. Acad. Sci.*, **111**, 3228–3232, doi:10.1073/pnas.1312330110.

Watanabe, M., and Coauthors, 2010: Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, **23**, 6312–6335, doi: 10.1175/2010JCLI3679.1

Wehner, M., D. R. Easterling, J. H. Lawrimore, R. R. Heim, R. S. Vose, and B. D. Santer, 2011: Projections of Future Drought in the Continental United States and Mexico. *J. Hydrometeor.*, **12**, 1359–1377, doi:10.1175/2011JHM1351.1.

World Bank, 2012: *Food Price Watch*, August 2012, World Bank, Washington, DC, http://documents.worldbank.org/curated/en/2012/08/16750795/food-price-watch.

Wuebbles, D., and Coauthors, 2014: CMIP5 Climate Model Analyses: Climate Extremes in the United States. *Bull. Amer. Meteor. Soc.*, **95**, 571–583, doi:10.1175/BAMS-D-12-00172.1.

Wu, R., and J. L. Kinter, 2009: Analysis of the Relationship of U.S. Droughts with SST and Soil Moisture: Distinguishing the Time Scale of Droughts. *J. Climate*, **22**, 4520–4538, doi:10.1175/2009JCLI2841.1.

Xia, Y., and Coauthors, 2012a: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, doi:10.1029/2011JD016048.

——, and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.*, **117**, D03110, doi:10.1029/2011JD016051.

——, and Coauthors, 2013: Validation of Noah-Simulated Soil Temperature in the North American Land Data Assimilation System Phase 2. *J. Appl. Meteor. Climatol.*, **52**, 455–471, doi:10.1175/JAMC-D-12-033.1.

——, J. Sheffield, M. B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood, 2014: Evaluation of multi-model simulated soil moisture in NLDAS-2. *J. Hydrol.*, **512**, 107–125, doi:10.1016/j.jhydrol.2014.02.027.

Xin, X., T. Wu, and J. Zhang, 2013: Introduction of CMIP5 simulations carried out with the climate system models of Beijing Climate Center (in Chinese). *Adv. Climate Change Res.*, **4**, 41–49, doi: 10.3724/SP.J.1248.2013.041.

Yukimoto, S., and Coauthors, 2012: A new global climate model of the Meteorological Research Institute: MRI-CGCM3— Model description and basic performance. *J. Meteor. Soc. Japan*, **90A**, 23–64, doi:10.2151/jmsj.2012-A02.

Zhang, Z. S., and Coauthors, 2012: Pre-industrial and mid-Pliocene simulations with NorESM-L. *Geosci. Model Dev.*, **5**, 523–533, doi:10.5194/gmd-5-523-2012.

TABLES

Model	Center	Type	Atmospheric Horizontal Resolution (lon. x lat.)	Number of model levels	Reference
ACCESS1-0/3	Commonwealth Scientific and Industrial Research Organization/Bureau of Meteorology, Australia	AO	1.875 x 1.25	38	Bi et al. (2012)
BCC-CSM1.1	Beijing Climate Center, China Meteorological Administration, China	ESM	2.8 x 2.8	26	Xin et al. (2012)
BCC-CSM1.1-M	Beijing Climate Center, China Meteorological Administration, China	ESM	1.125 x 1.125	26	Xin et al. (2012)
CanESM2	Canadian Center for Climate Modeling and Analysis, Canada	ESM	2.8 x 2.8	35	Arora et al. (2011)
CMCC-CM	Centro Euro-Mediterraneo sui Cambiamenti Climatici Climate Model, Italy	AO	0.75 x 0.75	31	Scoccimarro et al. (2011)
CNRM-CM5.1	National Centre for Meteorological Research, France	AO	1.4 x 1.4	31	Voldoire et al. (2013)
FGOALS-S2.0	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	AO	2.8 x 1.6	26	Bao et al. (2012)
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, USA	AO	2.5 x 2.0	48	Donner et al. (2011)
GFDL-ESM2G/M	NOAA Geophysical Fluid Dynamics Laboratory, USA	ESM	2.5 x 2.0	48	Donner et al. (2011)
HADGEM2-CC	Met Office Hadley Centre, UK	ESM	1.875 x 1.25	60	Jones et al. (2011)
INMCM4	Institute for Numerical Mathematics, Russia	AO	2.0 x 1.5	21	Volodin et al. (2010)

IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	ChemESM	3.75 x 1.9	39	Dufresne et al. (2012)
IPSL-CM5A-MR	Institut Pierre Simon Laplace, France	ChemESM	2.5 x 1.25	39	Dufresne et al. (2012)
IPSL-CM5B-LR	Institut Pierre Simon Laplace, France	ChemESM	3.75 x 1.9	39	Dufresne et al. (2012)
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	AO	1.4 x 1.4	40	Watanabe et al. (2010)
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	ESM	2.8 x 2.8	80	Watanabe et al. (2010)
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	ChemESM	2.8 x 2.8	80	Watanabe et al. (2010)
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	ESM	1.9 x 1.9	47	Giorgetta et al. (2013)
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany	ESM	1.9 x 1.9	47	Giorgetta et al. (2013)
MRI-CGCM3	Meteorological Research Institute, Japan	AO	1.125 x 1.121	48	Yukimoto et al. (2011)
MRI-ESM1	Meteorological Research Institute, Japan	ESM	1.125 x 1.121	48	Yukimoto et al. (2011)
NorESM1-M	Norwegian Climate Center, Norway	ESM	2.5 x 1.9	26	Zhang et al. (2012)

Table 1. CMIP5 models evaluated in this study and their attributes. Model types are: Atmosphere-Ocean coupled (AO), Earth System Model (ESM), and Earth System Model Chemistry coupled (ChemESM)

Statistic	Definition
Yearly Frequency	$F_Y = \frac{n_{extremes}}{n_{years}}$
Mean Duration	$\bar{D} = \frac{1}{n_{years} n_{events}} \sum_{i=1}^{n_{years}} \sum_{j=1}^{n_{events}} d_{i,j}$
Mean Intensity	Droughts: $\bar{I} = \frac{1}{n_{years} n_{events} n_{months}} \sum_{i=1}^{n_{years}} \sum_{j=1}^{n_{events}} \sum_{k=1}^{n_{months}} (1 - p_{i,j,k})$
	Heat Waves: $\bar{I} = \frac{1}{n_{years} n_{events} n_{days}} \sum_{i=1}^{n_{years}} \sum_{j=1}^{n_{events}} \sum_{k=1}^{n_{days}} (1 - p_{i,j,k})$
Mean Severity	Droughts: $\bar{S} = \frac{1}{n_{years} n_{events}} \sum_{i=1}^{n_{years}} \sum_{j=1}^{n_{events}} \sum_{k=1}^{n_{months}} (1 - p_{i,j,k})$
	Heat Waves: $\bar{S} = \frac{1}{n_{years} n_{events}} \sum_{i=1}^{n_{years}} \sum_{j=1}^{n_{events}} \sum_{k=1}^{n_{days}} p_{i,j,k}$

Table 2. Definitions of the statistics used to characterize droughts and heat waves. Here, $n_{extremes}$ is the number of years with at least one drought or heat wave, n_{years} is the number of years in the historical or future period, n_{events} is the number of events per season, $d_{i,j}$ is the duration of event *j* in year *i*, n_{months} and n_{days} are the number of months and days a drought and a heat wave lasted, respectively, and $p_{i,j,k}$ is the percentile during month/day *k* in event *j* in year *i*. Note that for droughts the intensity is defined as 1-$p_{i,j,k}$, such that drier conditions with lower percentiles translate to a higher intensity value. This definition of severity combines the duration of each event as well as the deviations from the threshold.

Model	Crop Area						Crop Upper						Crop Lower					
	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi
ACCESS1-0	22.3	16.3	-0.9	27.6	1.4	0.11	20.5	24.2	2.9	40.0	1.9	0.31	25.3	11.7	-4.9	14.7	0.9	-0.07
ACCESS1-3	26.8	19.0	-1.5	34.9	2.3	0.14	26.2	37.0	-9.1	43.7	3.0	0.35	28.2	6.1	0.7	24.3	1.8	-0.01
bcc-csm1-1	-1.9	-16.5	-26.2	-12.6	2.0	0.06	16.9	1.4	-13.3	8.7	0.8	0.22	-16.6	-31.4	-37.9	-35.9	2.7	-0.07
bcc-csm1-1-m	6.8	-3.7	-51.6	-23.3	3.9	-0.01	19.0	17.4	-45.2	-2.1	2.9	0.22	-3.4	-20.5	-60.0	-48.8	4.7	-0.23
CanESM2	4.7	0.0	-41.3	-15.7	6.0	0.13	13.8	12.3	-46.1	-12.1	7.6	0.25	-4.5	-9.0	-39.5	-20.6	4.6	0.02
CMCC-CM	12.8	14.2	20.6	33.7	-1.7	0.03	16.3	22.9	31.8	34.4	-2.0	0.13	10.6	7.7	10.1	30.6	-1.4	-0.09
CNRM-CM5	-4.3	10.1	-17.8	3.5	0.7	0.35	4.5	5.9	-24.3	-1.4	1.2	0.47	-10.5	15.7	-14.0	5.5	0.5	0.24
FGOALS-g2	0.7	-9.2	-15.3	-6.9	-1.6	0.05	11.5	1.5	-17.6	5.9	-2.4	0.15	-8.2	-18.4	-16.1	-22.7	-1.3	-0.04
GFDL-CM3	9.2	11.6	5.9	21.1	-1.6	0.09	20.8	18.2	7.7	24.9	-1.7	0.21	-0.2	6.6	2.4	17.5	-2.0	-0.07
GFDL-ESM2G	-2.3	-1.9	4.3	25.7	-0.4	0.08	17.1	12.1	9.3	33.7	-0.8	0.23	-16.7	-11.9	1.2	18.3	-0.7	-0.06
GFDL-ESM2M	2.2	8.5	-2.2	20.1	0.2	0.10	19.6	17.3	9.8	30.6	0.0	0.20	-10.8	2.1	-10.3	12.6	-0.2	0.01
HadGEM2-CC	9.8	15.0	-13.9	21.2	1.1	0.20	6.8	16.1	-17.6	27.3	2.1	0.33	12.6	13.4	-13.3	11.4	0.3	0.08
inmcm4	6.6	23.6	-8.2	25.3	-1.3	0.08	20.9	29.1	-3.3	41.9	-0.9	0.16	-3.9	21.7	-16.2	6.3	-1.8	0.04
IPSL-CM5A-LR	1.6	-19.9	0.2	10.7	0.3	0.15	23.7	5.0	-10.3	11.2	0.3	0.20	-15.4	-39.0	5.4	6.5	-0.2	0.14
IPSL-CM5A-MR	11.1	-22.4	-18.8	6.1	1.7	0.10	31.9	5.4	-23.6	8.2	1.8	0.22	-5.3	-45.7	-18.9	-0.6	1.4	0.01
IPSL-CM5B-LR	8.6	-6.8	13.0	6.0	-1.4	0.02	22.3	-4.6	-10.4	-0.7	-1.6	0.18	-3.0	-11.4	26.3	6.8	-1.6	-0.09
MIROC5	3.1	-0.4	-3.9	16.6	2.6	0.07	16.7	0.3	-1.9	14.9	3.3	0.27	-8.1	-0.2	-5.2	20.2	2.0	-0.09
MIROC-ESM	-13.6	-3.8	-18.8	10.0	3.2	-0.03	10.4	3.5	-11.9	20.7	4.2	0.21	-30.8	-7.3	-26.2	2.2	2.5	-0.25
MIROC-ESM-CHEM	-12.4	-6.4	-15.1	9.4	3.1	-0.02	9.1	1.6	-8.6	19.5	4.1	0.16	-28.3	-10.6	-20.0	3.0	2.3	-0.15
MPI-ESM-LR	10.2	22.2	21.0	32.8	-0.5	0.13	27.3	34.3	25.0	34.0	-0.9	0.31	0.2	15.1	19.3	29.1	0.0	-0.03
MPI-ESM-MR	3.0	18.2	17.1	32.4	-0.1	0.19	27.7	33.0	22.0	36.4	-0.5	0.35	-14.4	8.5	13.4	26.0	0.3	0.05
MRI-CGCM3	11.7	17.5	5.2	25.5	-0.7	-0.32	15.8	21.2	11.9	27.9	-0.6	-0.04	7.9	13.5	-1.6	18.6	-1.0	-0.55
MRI-ESM1	13.4	14.1	4.7	25.1	-0.6	-0.35	21.0	17.2	9.7	27.4	-0.5	-0.12	6.6	12.2	-0.6	18.4	-0.9	-0.58
NorESM1-M	-17.4	-8.3	10.9	37.6	-0.7	-0.12	3.1	0.3	10.7	45.2	-0.3	0.08	-33.9	-15.4	12.7	30.4	-1.3	-0.30

Table 3. Biases in DJF Prcp (mm/month), MAM Prcp (mm/month), JJA Prcp (mm/month), JJA ET (mm/month), JJA Tas (K), JJA ϕ (unitless) for the Crop sub-regions.

Model	Crop Area						Crop Upper						Crop Lower					
	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi	DJF Prcp	MAM Prcp	JJA Prcp	JJA ET	JJAT Tas	JJA Phi
ACCESS1-0	10.5	5.6	-25.9	-22.1	6.8	-0.15	13.2	13.	-30.4	-19.6	7.2	-0.19	8.2	-1.5	-23.2	-27.4	6.6	-0.14
ACCESS1-3	20.7	20.5	1.8	6.1	4.8	0.10	18.2	24.	0.7	5.4	5.2	0.11	23.3	18.6	3.9	6.5	4.5	0.07
bcc-csm1-1	0.9	1.9	-8.4	-0.1	4.8	-0.05	5.8	12.	2.6	6.4	4.7	0.03	-3.0	-3.9	-14.2	-5.3	5.0	-0.11
bcc-csm1-1-m	4.1	5.4	-6.2	-0.6	4.7	-0.02	8.8	10.	-6.9	1.2	4.8	-0.02	1.4	1.2	-5.5	-3.4	4.7	-0.03
CanESM2	9.2	3.3	0.4	-0.5	5.9	-0.03	13.1	14.	0.0	1.0	6.4	-0.01	7.3	-3.2	2.6	-0.5	5.5	-0.07
CMCC-CM	4.1	9.5	16.3	11.4	5.2	0.18	10.7	19.	28.3	18.0	4.7	0.07	-0.4	3.0	6.8	8.1	5.6	0.24
CNRM-CM5	3.0	11.7	0.8	6.0	4.4	-0.02	7.2	18.	-0.9	8.4	4.8	0.00	0.4	6.9	2.1	3.8	4.1	-0.04
FGOALS-g2	1.3	-3.4	-7.7	0.8	4.8	0.01	12.5	5.5	2.8	8.3	5.1	0.08	-5.4	-9.4	-15.0	-5.2	4.8	-0.05
GFDL-CM3	18.6	24.1	8.3	21.5	7.2	0.07	18.6	32.	7.3	25.8	8.2	0.13	20.6	20.4	8.8	18.1	6.4	0.06
GFDL-ESM2G	2.9	9.8	0.7	9.6	3.9	0.00	7.7	22.	11.6	19.7	3.6	-0.02	-0.4	2.0	-7.9	2.3	4.2	0.03
GFDL-ESM2M	3.4	5.5	-1.1	4.9	3.9	0.03	10.5	16.	3.9	13.1	3.6	0.02	-0.9	-0.5	-5.9	-1.2	4.2	0.05
HadGEM2-CC	23.1	5.9	-26.7	-17.2	8.1	-0.16	17.7	13.	-20.9	-8.3	8.5	-0.11	27.4	2.3	-32.6	-28.3	7.8	-0.23
inmcm4	7.6	17.0	-2.5	-0.9	3.2	-0.02	11.9	24.	-4.2	2.1	2.9	-0.05	5.4	11.5	-1.6	-4.1	3.5	-0.01
IPSL-CM5A-LR	-9.4	-3.2	-5.7	-0.2	5.8	0.04	0.0	2.4	-8.7	1.5	6.1	0.09	-16.2	-6.9	-3.7	-1.3	5.6	-0.02
IPSL-CM5A-MR	-21.2	-0.8	-14.5	-9.6	6.5	-0.02	-10.1	6.8	-13.1	-4.4	6.7	0.04	-29.6	-4.8	-16.5	-14.0	6.5	-0.09
IPSL-CM5B-LR	11.7	8.0	-1.9	8.6	4.3	0.07	10.1	6.2	-0.9	8.1	4.9	-0.04	12.8	10.9	-0.1	9.8	4.0	0.13
MIROC5	3.2	6.9	-10.6	-0.2	5.4	0.18	6.6	13.	-15.9	2.2	5.9	0.14	0.9	1.0	-7.6	-2.4	5.1	0.23
MIROC-ESM	6.6	16.0	-12.6	-1.6	7.2	0.04	8.1	23.	-13.6	3.9	7.4	0.00	4.7	10.9	-12.4	-7.8	7.2	0.11
MIROC-ESM-CHEM	5.1	16.1	-10.2	-0.6	7.2	0.03	9.0	23.	-12.1	5.5	7.4	0.08	2.8	11.2	-10.7	-8.2	7.2	-0.01
MPI-ESM-LR	8.9	16.2	6.1	9.2	5.3	0.07	9.6	24.	7.8	11.4	5.8	0.05	8.5	10.4	3.2	7.0	5.0	0.06
MPI-ESM-MR	17.4	23.2	7.0	10.7	4.9	0.10	10.4	19.	4.7	7.8	5.3	0.08	25.1	27.3	9.3	13.3	4.6	0.11
MRI-CGCM3	13.6	14.7	9.1	14.5	3.4	0.04	18.2	13.	10.9	18.1	3.6	-0.08	11.1	17.2	8.7	10.9	3.3	0.11
MRI-ESM1	12.9	17.2	16.0	16.4	3.5	0.06	11.2	22.	18.5	20.4	3.7	0.01	15.1	12.2	13.6	12.2	3.3	0.07
NorESM1-M	5.8	9.8	-7.6	0.3	5.0	0.12	9.4	11.	-6.4	1.9	5.7	0.08	3.6	8.6	-7.9	-1.2	4.5	0.17

Table 4. Projected changes in DJF Prcp (mm/month), MAM Prcp (mm/month), JJA Prcp (mm/month), JJA ET (mm/month), JJA Tas (K), JJA ϕ (unitless) for the Crop sub-regions under RCP 8.5 between 1979-2005 and 2070-2099.

FIGURE CAPTIONS

Figure 1. Scatter plots of JJA mean ET versus JJA interannual standard deviation of ET for two land-surface models in NLDAS-2: a) VIC and b) Noah. Each point represents a grid cell. Grid-cells are labeled as water-limited if they have a significant ($p<0.05$) correlation between SM and ET larger than 0.3, as radiation-limited if they have a significant negative correlation with magnitude larger than 0.3, and as transitional otherwise. This shows that water-limited regions can have low variability and high mean, as well as low mean and high variability.

Figure 2. Maps of two coupling metrics calculated for JJA (1979-2005) from NLDAS-2 ensemble means (i.e. VIC and Noah). Metric γ in panel (a) is a normalized version of Dirmeyer et al. (2013b) from Equation 1. Metric ϕ in panel (b) results from Equation 2 and it is the one used in the rest of this study. The sub-regions used in this study are defined based on patterns in ϕ.

Figure 3. Distributions (a-f) and cross-correlations across 24 CMIP5 models (g-n) of historical biases in climatologies of MAM and JJA Prcp, JJA ET, JJA Tas, JJA ϕ, and JJA correlation between SM and ET compared to the climatologies from NLDAS-2 during 1979-2005. Black diamonds in (a-f) represent the NLDAS-2 ensemble means. Hatches in (g-n) represent statistically significant ($p<0.05$) correlations.

Figure 4. Distributions (a-f) and cross-correlations across 24 CMIP5 models (g-n) of future projections normalized by mean global change in JJA near-surface air temperature

for the same variables as in Figure 3. Hatches in (g-n) represent statistically significant ($p<0.05$) correlations.

Figure 5. Relationship between historic biases in JJA precipitation and future hydro-climatic changes. Slopes (left panels) and y-intercepts (right panels) of the linear regressions fitted between historical biases in JJA Prcp and normalized future changes in MAM and JJA Prcp, JJA ET, JJA Tas, JJA ρ(SM, ET), and JJA ϕ for each sub-region. Error bars represent the standard errors and the hatch represents statistical significant values ($p<0.05$). The numbers above each bar in the left panels represent the proportion of the variance explained by each relationship.

Figure 6. Projected percentage changes in drought yearly frequency and severity for each climate model and sub-region, under RCP8.5.

Figure 7. Projected percentage changes in heat wave yearly frequency and severity for each climate model and sub-region, under RCP8.5.

Figure 8. Spearman rank correlations between historical biases in JJA Prcp and the absolute projected changes in drought characteristics (a-d), and biases in JJA Tas and heat wave characteristics (e-h) across models for each region. Hatched bars represent statistically significant results ($p<0.05$).

Figure 9. Interquartile ranges of future changes for characteristics of droughts and heat waves over the Crop Area. The blue line represent the median range when randomly sampling a subset of models, and the envelope corresponds to the 10th and 90th percentiles (derived from 1,000 repetitions for each subset). Colored lines represents the range of uncertainty when using the models along the order of the rankings based on JJA precipitation (red), JJA evapotranspiration (orange), JJA near-surface air temperature (teal), and JJA land-atmospheric coupling (indigo).

FIGURES

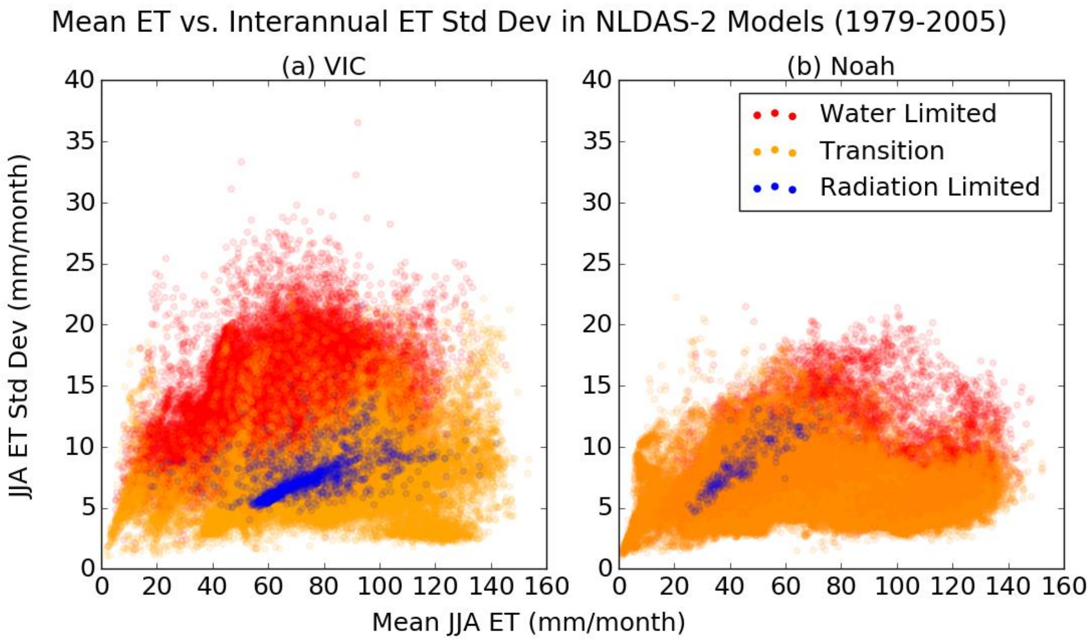

Figure 1. Scatter plots of JJA mean ET versus JJA interannual standard deviation of ET for two land-surface models in NLDAS-2: (a) VIC and (b) Noah. Each point represents a grid-cell. Grid-cells are labeled as water-limited if they have a significant ($p<0.05$) correlation between SM and ET larger than 0.3, as radiation-limited if they have a significant negative correlation with magnitude larger than 0.3, and as transitional otherwise. This shows that water-limited regions can have low variability and high mean, as well as low mean and high variability.

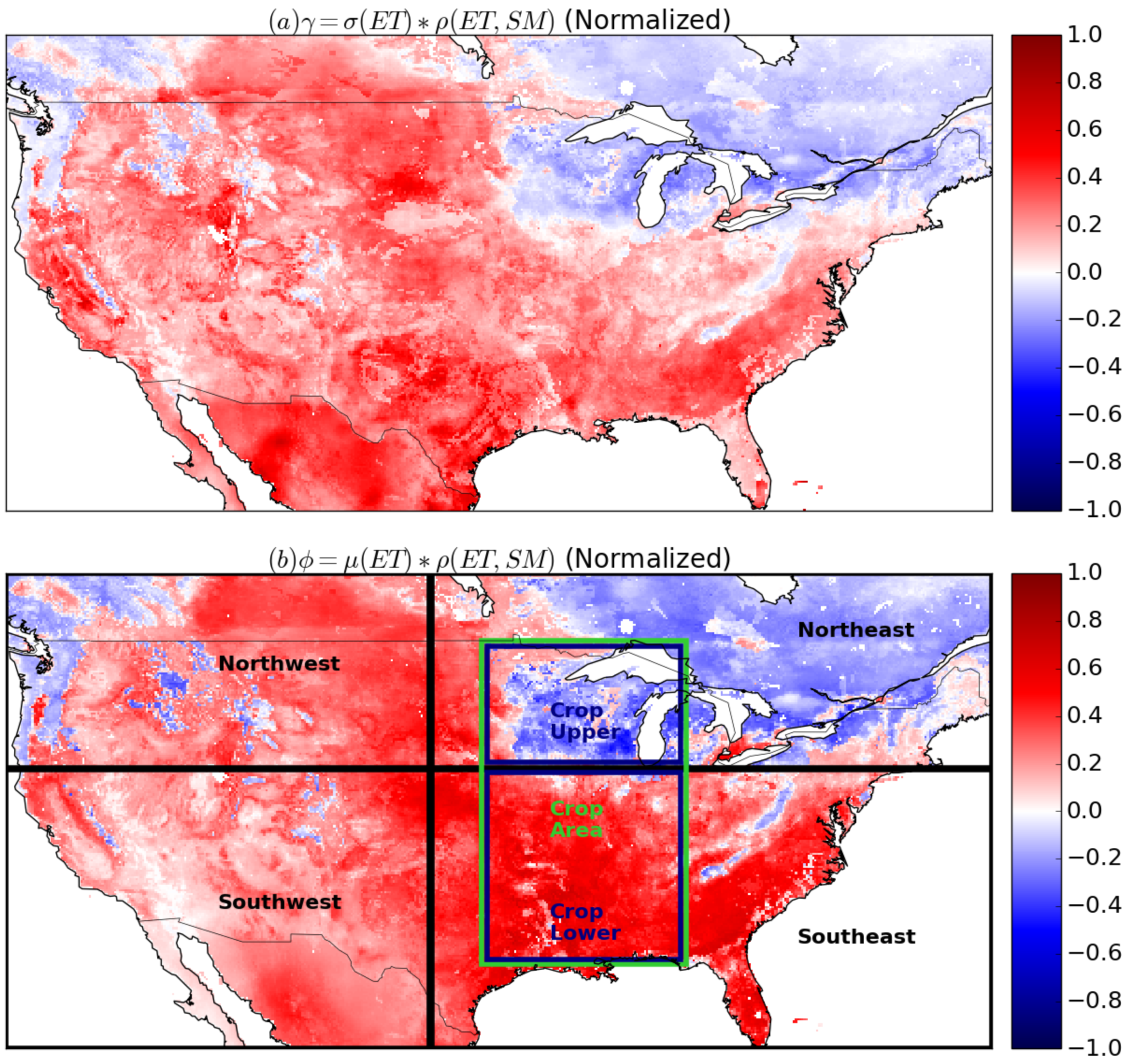

Figure 2. Maps of two coupling metrics calculated for JJA (1979-2005) from NLDAS-2 ensemble means (i.e. VIC and Noah). Metric γ in panel (a) is a normalized version of Dirmeyer et al. (2013b) from Equation 1. Metric ϕ in panel (b) results from Equation 2 and it is the one used in the rest of this study. The sub-regions used in this study are defined based on patterns in ϕ.

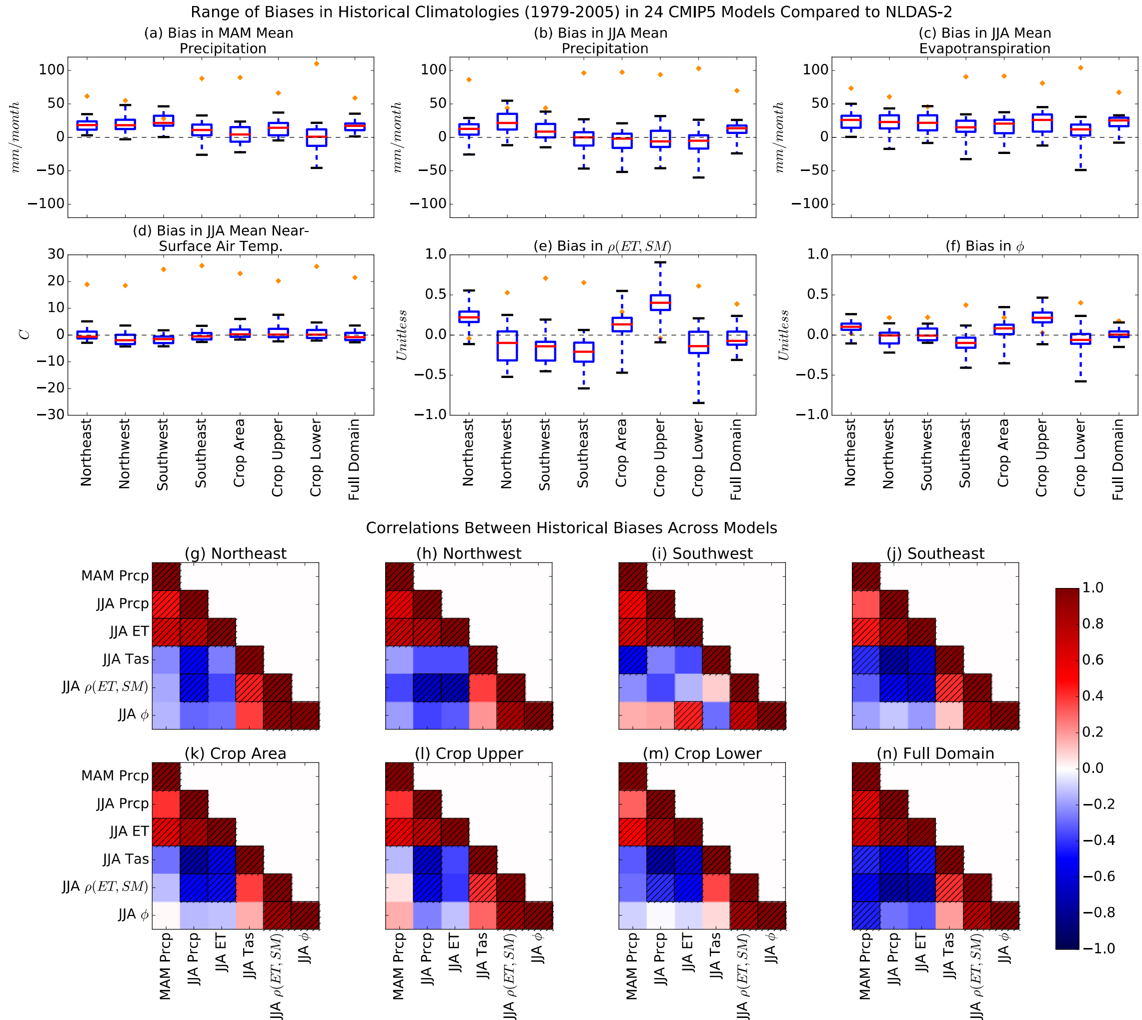

Figure 3. Distributions (a-f) and cross-correlations across 24 CMIP5 models (g-n) of historical biases in climatologies of MAM and JJA Prcp, JJA ET, JJA Tas, JJA ϕ, and JJA correlation between SM and ET compared to the climatologies from NLDAS-2 during 1979-2005. Orange diamonds in (a-f) represent the NLDAS-2 ensemble means. Hatches in (g-n) represent statistically significant ($p<0.05$) correlations.

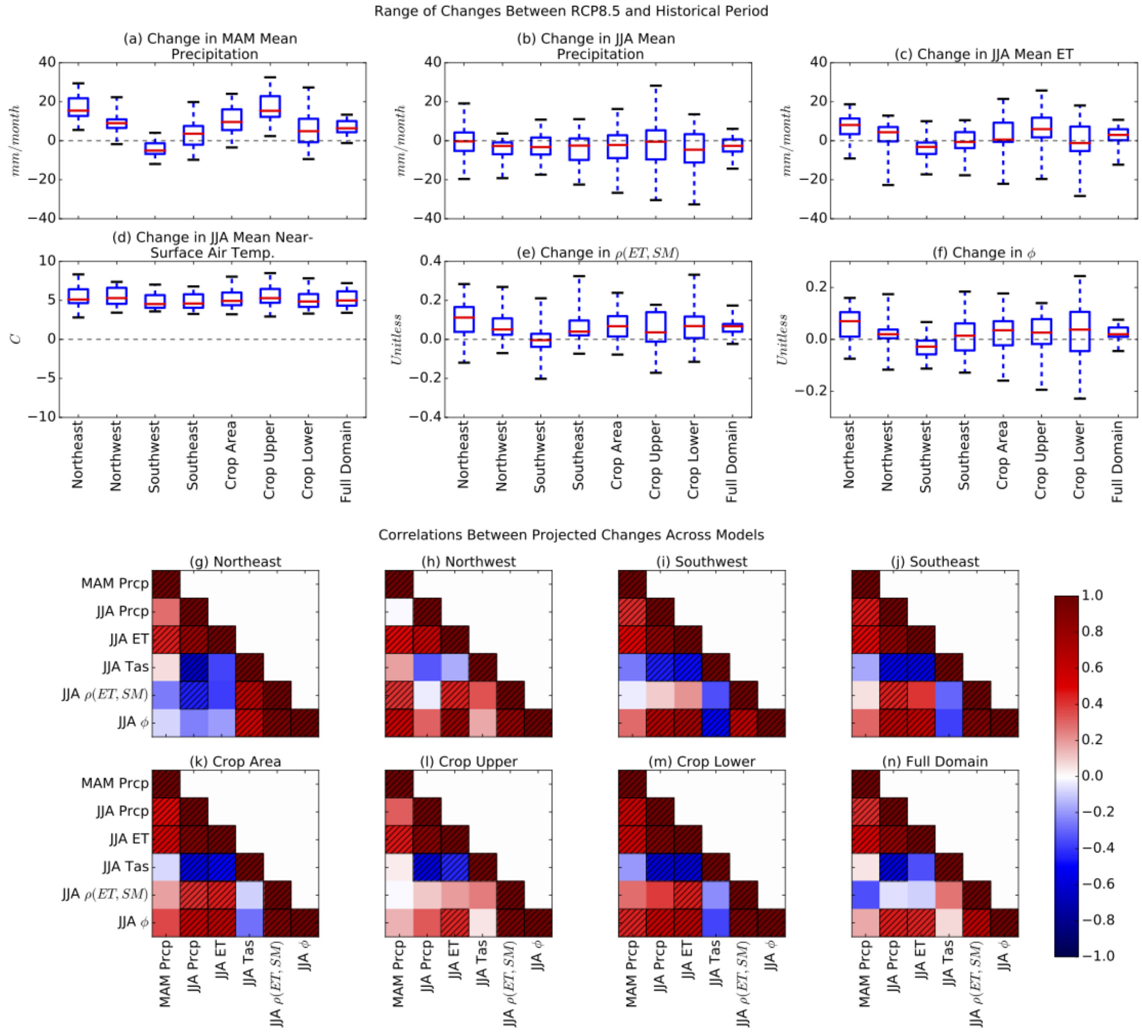

Figure 4. Distributions (a-f) and cross-correlations across 24 CMIP5 models (g-n) of future projections normalized by mean global change in JJA near-surface air temperature for the same variables as in Figure 3. Hatches in (g-n) represent statistically significant ($p<0.05$) correlations.

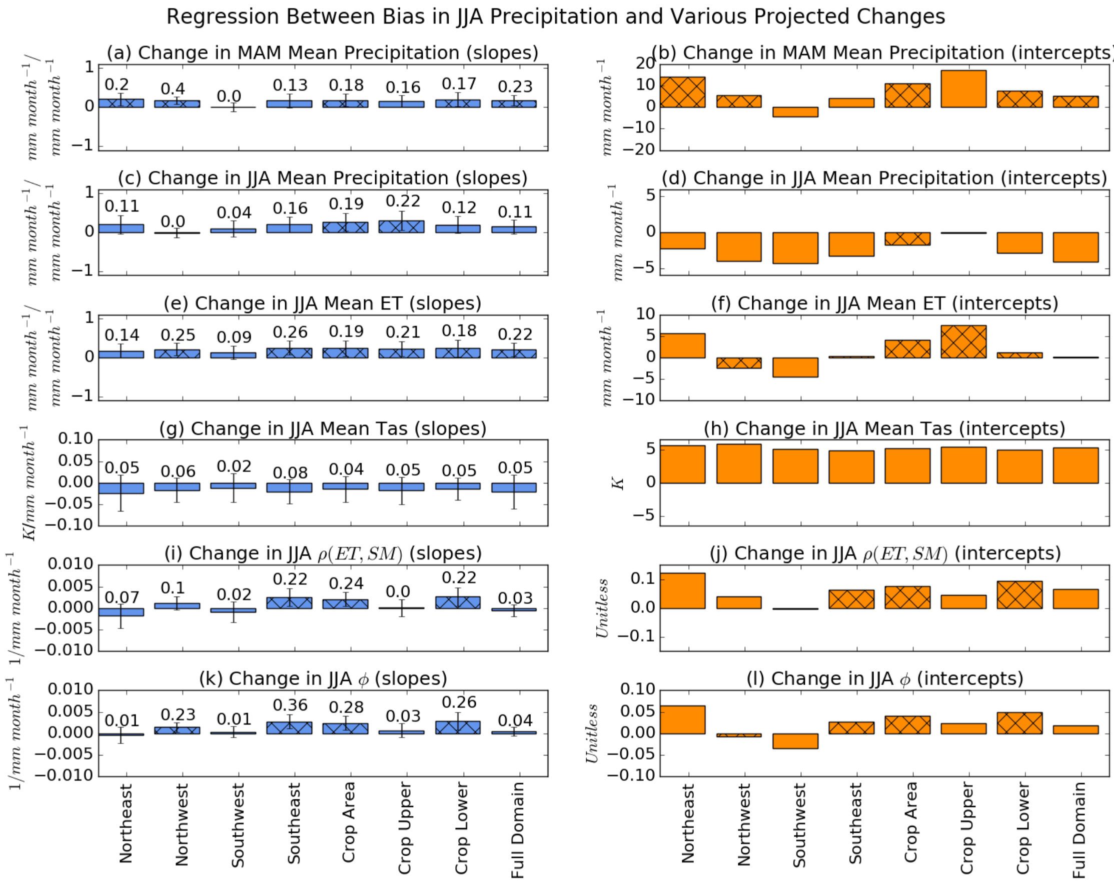

Figure 5. Relationship between historic biases in JJA precipitation and future hydro-climatic changes. Slopes (left panels) and y-intercepts (right panels) of the linear regressions fitted between historical biases in JJA Prcp and normalized future changes in MAM and JJA Prcp, JJA ET, JJA Tas, JJA ρ(SM, ET), and JJA ϕ for each sub-region. Error bars represent the standard errors and the hatch represents statistical significant values (p<0.05). The numbers above each bar in the left panels represent the proportion of the variance explained by each relationship.

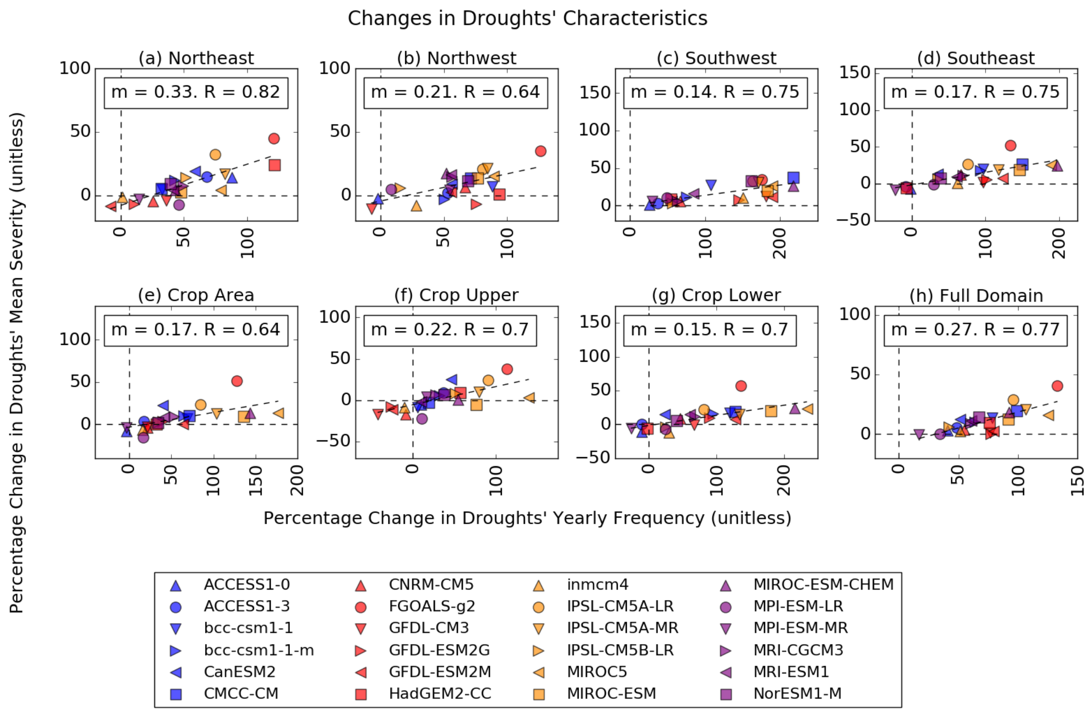

Figure 6. Projected percentage changes in drought yearly frequency and severity for each climate model and sub-region, under RCP8.5.

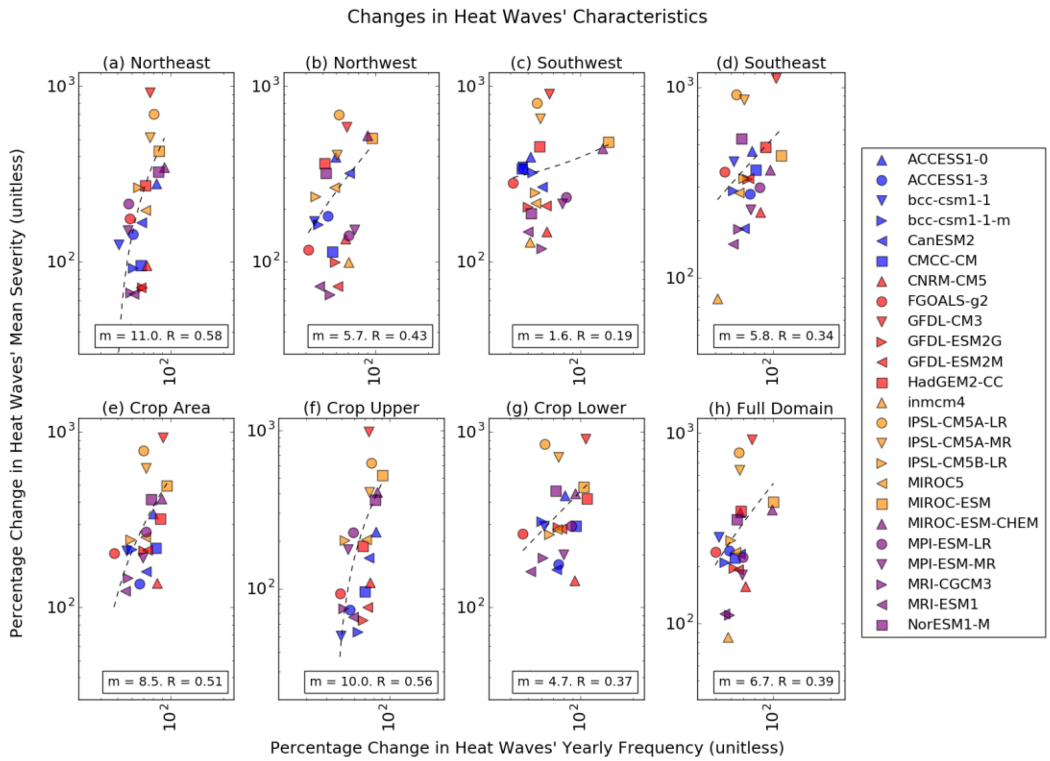

Figure 7. Projected percentage changes in heat wave yearly frequency and severity for each climate model and sub-region, under RCP8.5.

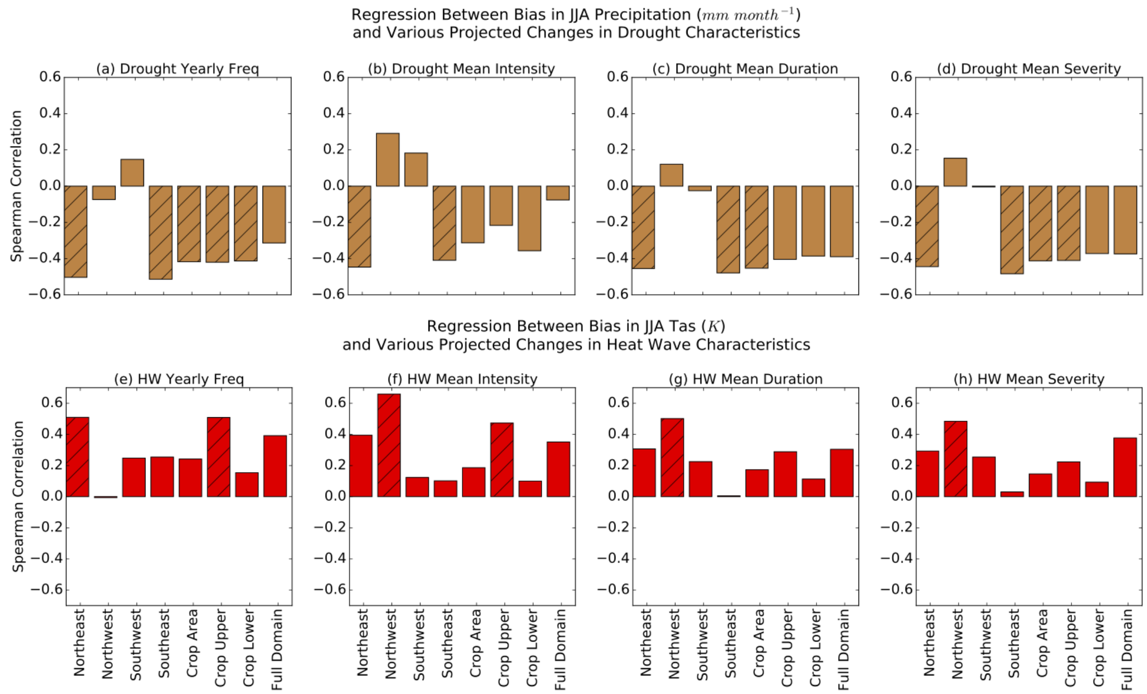

Figure 8. Spearman rank correlations between historical biases in JJA Prcp and the absolute projected changes in drought characteristics (a-d), and biases in JJA Tas and heat wave characteristics (e-h) across models for each region. Hatched bars represent statistically significant results ($p<0.05$).

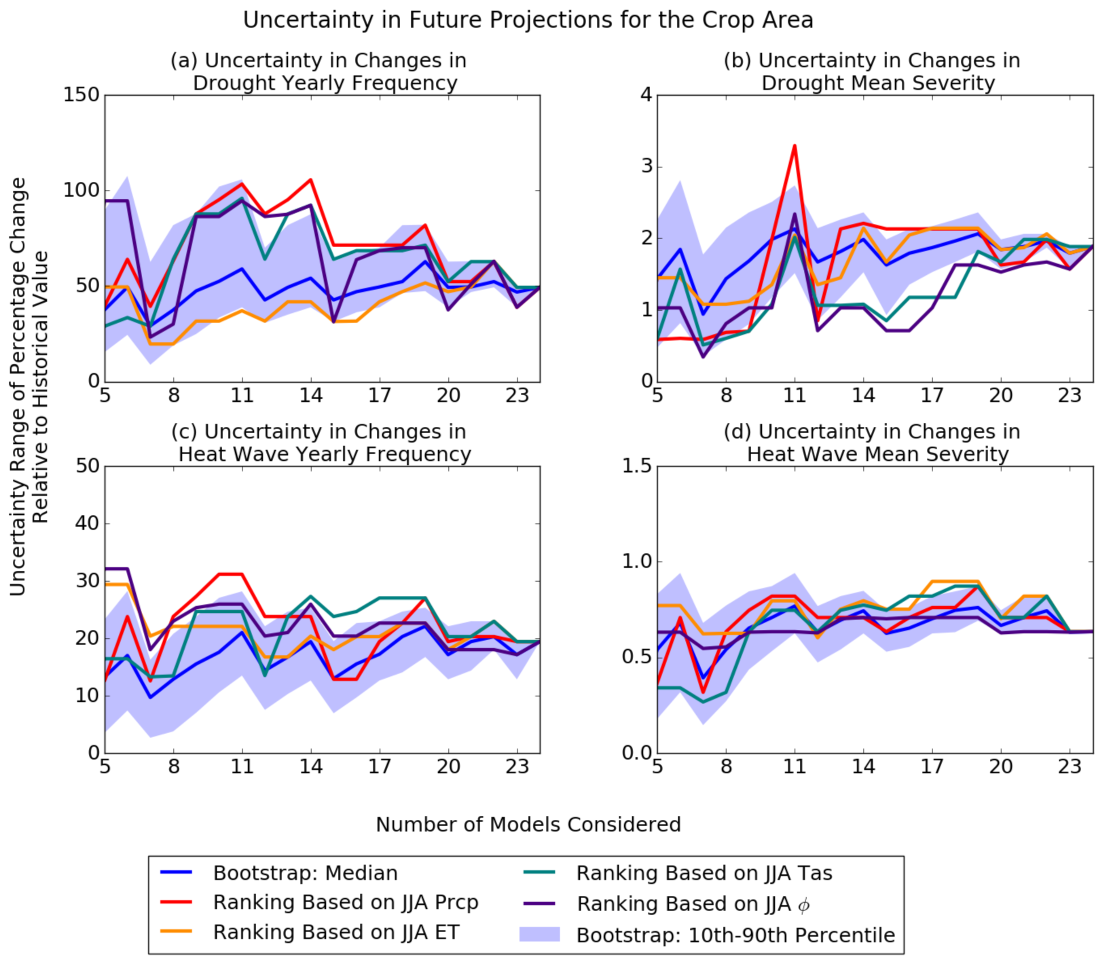

Figure 9. Interquartile ranges of future changes for characteristics of droughts and heat waves over the Crop Area. The blue line represent the median range when randomly sampling a subset of models, and the envelope corresponds to the 10th and 90th percentiles (derived from 1,000 repetitions for each subset). Colored lines represents the range of uncertainty when using the models along the order of the rankings based on JJA precipitation (red), JJA evapotranspiration (orange), JJA near-surface air temperature (teal), and JJA land-atmospheric coupling (indigo).