An Evaluation of Modeled Evaporation Regimes in Europe Using Observed Dry Spell Land Surface Temperature

PHIL P. HARRIS

Centre for Ecology and Hydrology, and National Centre for Earth Observation, Wallingford, United Kingdom

SONJA S. FOLWELL AND BELEN GALLEGO-ELVIRA

Centre for Ecology and Hydrology, Wallingford, United Kingdom

JOSÉ RODRÍGUEZ AND SEAN MILTON

Met Office, Exeter, United Kingdom

CHRISTOPHER M. TAYLOR

Centre for Ecology and Hydrology, and National Centre for Earth Observation, Wallingford, United Kingdom

(Manuscript received 15 September 2016, in final form 13 January 2017)

ABSTRACT

Soil moisture availability exerts control over the land surface energy partition in parts of Europe. However, determining the strength and variability of this control is impeded by the lack of reliable evaporation observations at the continental scale. This makes it difficult to refine the broad range of soil moisture–evaporation behaviors across global climate models (GCMs). Previous studies show that satellite observations of land surface temperature (LST) during rain-free dry spells can be used to diagnose evaporation regimes at the GCM gridbox scale. This relative warming rate (RWR) diagnostic quantifies the increase in dry spell LST relative to air temperature and is used here to evaluate a land surface model (JULES) both offline and coupled to a GCM (HadGEM3-A). It is shown that RWR can be calculated using outputs from an atmospheric GCM provided the satellite clear-sky sampling bias is incorporated. Both offline JULES and HadGEM3-A reproduce the observed seasonal and regional RWR variations, but with weak springtime RWRs in central Europe. This coincides with sustained bare soil evaporation (Ebs) during dry spells, reflecting previous site-level JULES studies in Europe. To assess whether RWR can discriminate between surface descriptions, the bare soil surface conductance and the vegetation root profile are revised to limit Ebs. This increases RWR by increasing the occurrence of soil moisture–limited dry spells, yielding more realistic springtime RWRs as a function of antecedent precipitation but poorer relationships in summer. This study demonstrates the potential for using satellite LST to assess evaporation regimes in climate models.

1. Introduction

The extreme 2003 European heat wave (Schär et al. 2004) has focused attention on the processes relating to high spring and summer temperatures in this region. While that particular event may have been forced partly by remote sea surface temperatures (Feudale and Shukla 2007, 2011), land–atmosphere feedbacks are also required to explain the magnitude and persistence of the heat wave (Miralles et al. 2012, 2014). More generally, there is observational evidence relating European summer temperature extremes to soil moisture through precipitation (Hirschi et al. 2011; Mueller and Seneviratne 2012; Quesada et al. 2012).

The main soil moisture–air temperature coupling operates through its control over the partition of surface net radiation into latent and sensible heat flux, such that evaporation is constrained by either water availability or by radiation. Eddy covariance flux measurements indicate that, for summer evaporation, Europe is divided into two hydroclimatic regions: radiation limited in the north and soil moisture limited in the south (Teuling et al. 2009). However, the location and extent of the related transition zone is uncertain.

Corresponding author e-mail: P. P. Harris, ppha@ceh.ac.uk

DOI: 10.1175/JHM-D-16-0227.1

The role of these land surface processes in heat waves and warm temperature extremes is reflected in atmospheric models. For example, the simulation of features such as the strength, location, and duration of the 2003 heat wave is influenced by soil moisture state (Bisselink et al. 2011; Ferranti and Viterbo 2006; Fischer et al. 2007b; Weisheimer et al. 2011). Similar results have been shown in case studies of other historical European heat wave events (Fischer et al. 2007a; Vautard et al. 2007; Zampieri et al. 2009) and in the temperature extremes of multidecadal regional (Jaeger and Seneviratne 2011; Lorenz et al. 2012) and global simulations (Seneviratne et al. 2013). Most studies emphasize the role of soil moisture control over the surface energy partition, but there is modeling evidence for an additional, but weaker, enhancement of the 2003 event temperatures through reduced vegetation leaf area index (LAI; Lorenz et al. 2013; Stéfanon et al. 2012).

Despite the success of some RCMs in simulating specific heat wave events, there is little intermodel consistency in European warm temperature extremes even when the models are driven by the same synoptic forcing (Lhotka and Kyselý 2015; Vautard et al. 2013). This has been attributed in part to a wide variation in modeled land surface processes. For example, the RCMs in the ENSEMBLES project that have low mean evaporative fraction also tend to simulate high interannual summer temperature variability (Fischer et al. 2012), indicating different strengths of soil moisture limitation. Similarly, these RCMs exhibit different evaporative fraction changes between warm and cool summers (Stegehuis et al. 2013), ranging from no change to lower evaporative fraction in warm years. Indeed, these models disagree for much of central and western Europe whether interannual variability in summer evaporation is limited by radiation or soil moisture (Boé and Terray 2014), corresponding to disagreement in the location of the hydroclimatic transition zone. This ensemble spread in surface coupling is also a feature of the present-day climates of GCMs contributing to the CMIP3 (Boé and Terray 2008) and CMIP5 (Berg et al. 2015) ensembles.

This variation in modeled surface processes has implications beyond heat wave and seasonal predictability. Dirmeyer et al. (2013) show that there is little consensus across the CMIP5 ensemble about historical trends in the summer mean soil moisture or interannual variability in soil moisture and evaporation, despite agreement about declines in the mean available energy. Similarly, they describe how CMIP5 future projections show ensemble agreement in European summer mean soil moisture drying, stronger soil moisture–evaporation coupling, and increased sensible heat flux, but disagree on the sign of change in evaporation mean or variability.

The processes affecting present-day interannual summer climate variability in a particular model typically have a strong influence over the regional future climate change expressed by the model under anthropogenic global warming (Seneviratne et al. 2006). Specifically, models that exhibit large present-day summer temperature variability simulate little change or a reduction in that variability (Fischer et al. 2012). Such behavior could be symptomatic of a model frequently exhibiting severe soil moisture limitation in a changed climate. Errors in these processes may contribute to the finding that present-day model temperature biases are greater at warmer temperatures (Christensen et al. 2008). In a warming climate, therefore, regional climate change projections themselves are likely to have warm biases (Boberg and Christensen 2012; Christensen and Boberg 2012), but these biases may have an upper limit because of the physical limits on model drying (Bellprat et al. 2013). Nonetheless, while these biases are necessarily corrected in assessments of regional climate change, it is more desirable to reduce these errors by evaluating and improving model processes.

Evaluation of these surface processes in models would, ideally, involve comparing them directly with observations of surface sensible and latent heat flux and related quantities like evaporative fraction. In situ flux observations from eddy covariance provide useful information about observed (Teuling et al. 2006, 2009) and modeled (Blyth et al. 2010) surface processes at the site scale but are not representative of atmospheric model grid boxes with length scales of 50–200 km. Earth observation (EO) from satellites potentially provides the desired spatial coverage but is only able to observe surface states (e.g., skin temperature) and not turbulent fluxes. Gridded estimates of evaporation, therefore, have typically been hybrid products, using EO data as inputs to some form of model. For example, Jung et al. (2010) combined EO and flux tower data using a machine-learning technique to produce global monthly evaporation estimates at 0.5° resolution. Similarly, Miralles et al. (2011) used several EO products to drive and constrain a simple water budget model to produce estimates of global daily evaporation at 0.25° resolution. These and other evaporation estimates were assessed in the LandFlux-EVAL initiative (Jiménez et al. 2011; Mueller et al. 2011) and were found to contain significant mutual differences, limiting their use for model evaluation. Moreover, the hybrid models can also share approaches with the GCM land surface schemes that their outputs are used to assess. In a related effort to derive an evaporation benchmark from these estimates,

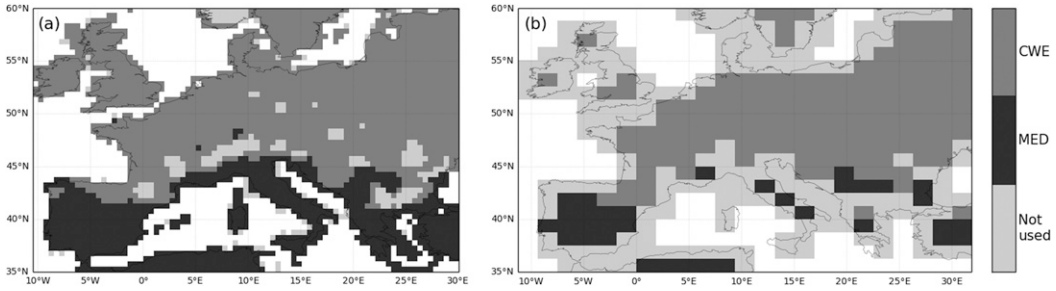

FIG. 1. European domains used for evaluation of (a) WFDEI/MODIS observations and offline JULES modeling and (b) global HadGEM3-A modeling. Regions used in the analysis are CWE and MED. In (b), the "not used" category includes both unused Köppen–Geiger classes and grid boxes that contain less than 90% land.

Mueller et al. (2013) produced synthesis datasets based on combinations of EO-derived data, offline land surface models, and reanalysis outputs.

Alternatively, we can use the well-observed satellite land surface temperature (LST) to provide indirect information about water limitation on the surface energy budget. Folwell et al. (2016) and Gallego-Elvira et al. (2016) described an LST-based diagnostic [relative warming rate (RWR)] that uses rain-free dry spells to infer the response of the surface energy partition to declining soil moisture. As the surface dries and evaporative fraction declines, LST increases more rapidly than the near-surface air temperature. Those studies demonstrated that seasonal and regional variations in RWR were consistent with a simple theoretical description of soil moisture constraints on evaporation. In this diagnostic, MODIS LST observations are composited across many dry spell events in order to determine a typical short-term surface response. Focusing on surface dry down can give clear signals because the land state is forced in one direction, and RWR is derived from quantities routinely available from GCMs so that the modeled large-scale behavior of the surface energy partition can be assessed.

Here, we use RWR for the first time to evaluate the behavior of a land surface model at large scales. Specifically, we assess the Joint UK Land Environment Simulator (JULES) land surface model for the European spring and summer, both offline and coupled to its host atmosphere model, HadGEM3-A. Offline JULES simulations at 0.5° are used to assess modeled RWR and its sensitivity to descriptions of land surface processes. This offline, composite diagnostic uses the historical dry spells and clear-sky sampling from the Water and Global Change (WATCH) Forcing Data–ERA-Interim (WFDEI) and MODIS data used to calculate observed RWR. We then describe how the observed clear-sky sampling can be reproduced from outputs of HadGEM3-A simulations and assess RWR in this model. Finally, we assess the sensitivity of HadGEM3-A RWR to the land surface process changes introduced in the offline JULES modeling.

In this paper, section 2 describes the data and models, and section 3 outlines the RWR calculation from Folwell et al. (2016) and describes the JULES modifications used in the offline and HadGEM3-A sensitivity simulations. The observed and HadGEM3-A-modeled dry spells are described in section 4. The dry spell composite RWR results for offline JULES and HadGEM3-A simulations are presented in sections 5 and 6, respectively.

2. Data and models

a. Region of study

In this paper, we concentrate on the same European domain that was used by Folwell et al. (2016). This section excludes most of Scandinavia and extends into North Africa (Fig. 1a). To aid interpretation of the results, we also split this European domain into two regions, central and western Europe (CWE) and Mediterranean (MED), which approximately correspond to the two hydroclimatic regions described by Teuling et al. (2009). These regions are based on Köppen–Geiger classifications (Peel et al. 2007), with MED comprising the classes cold semi-arid climate (BSk), hot summer Mediterranean climate (Csa), warm summer Mediterranean climate (Csb), humid subtropical climate (Cfa), cold desert climate (BWk), hot desert climate (BWh), hot semi-arid climate (BSh), and warm dry summer continental climate (Dsb), and CWE comprising the classes temperate maritime climate (Cfb) and warm summer humid continental climate (Dfb).

b. WFDEI surface meteorology

This study uses a gridded dataset of surface meteorology in order to 1) identify rain-free dry spells, 2) quantify variations in air temperature during dry

spells, and 3) force the JULES land surface model. The WFDEI (Weedon et al. 2014) is a global 0.5° gridded dataset of near-surface meteorology over land spanning the years 1979–2012. These data are based on 3-hourly diagnostics from ERA-Interim (Dee et al. 2011), which are interpolated spatially onto the regular 0.5° grid before bias corrections are applied.

The WFDEI precipitation data include bias corrections for monthly amounts and the number of wet days from the Climatic Research Unit (CRU) Time Series, using version 3.1 for the years 1979 to 2009 and version 3.21 for years after 2009 (CRU TS3.1 and CRU TS3.21, respectively). The number of wet days is corrected only in months when the equivalent monthly value in the spatially interpolated ERA-Interim data differs from CRU by more than 2 days, which is mainly in the tropics where there are fewer observations and reanalyses are less well constrained. The 2-m air temperature in WFDEI is derived from ERA-Interim 10-m air temperature, including a correction for elevation differences between the grids and bias correction to the CRU monthly mean and monthly diurnal temperature range. The surface incident shortwave radiation data include corrections for interannual variation in aerosol loading (ERA-Interim already accounts for the climatological variation), which is particularly important in the Northern Hemisphere during summer.

c. *MODIS land surface temperature*

The MODIS sensor provides LST observations on a sinusoidal grid of approximately 1 km resolution (level 3 product MOD11A1 Collection 5, accessed from https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a1). We use MODIS data from the polar-orbiting *Terra* platform for the 13-yr period 2000–12. This provides clear-sky daytime data once per day with an equatorial overpass time of approximately 1030 local time (LT). Data are only included for pixels with the best quality control flag (pixel-level quality assurance = 0) and with a view angle of less than 55° from nadir, as LST biases increase rapidly for greater angles (Trigo et al. 2008). We also exclude all data from 1-km pixels where the dominant IGBP land-cover class is wetland, land ice, water, or urban (classes 11, 13, 15, and 17), based on the 500-m MODIS Land-Cover Type Product (MCD12Q1).

Following Folwell et al. (2016), daily time series of MODIS LST on the 0.5° European domain in Fig. 1a are calculated using the mean of the available, cloud-free 1-km values in each grid box on each day. These aggregate data provide a more complete time series than is obtained from the 1-km data and at a scale more appropriate for comparison with global modeling. However, this aggregation introduces additional noise in LST from day-to-day variability in cloud cover affecting which portions of the grid box are sampled. We mitigate against gross sampling effects by rejecting 0.5° daily values derived from fewer than 100 pixels and by using temporal LST anomalies (LSTAs). These LSTA time series are constructed by first calculating a monthly mean LST climatology on the original 1-km MODIS grid using all years of cloud-free data. Daily anomalies are then calculated by sampling from this climatology based on the 1-km LST availability each day, with linear interpolation to day of month, before spatial averaging to the 0.5° grid.

d. *HadGEM3-A*

The Met Office Unified Model (MetUM; Walters et al. 2014) is a modeling framework used for both numerical weather prediction and climate simulation. Here, we use HadGEM3-A with science configuration Global Atmosphere 5.0 (GA5.0). This is a land–atmosphere configuration of the MetUM (version 8.3) with 85 vertical levels and a horizontal resolution of N96, which corresponds to a gridbox size of 1.85° × 1.25°. Simulations are forced with AMIP historical sea surface temperature boundary conditions from 1982 to 2008 on a 360-day year, which was run as two separate realizations for 1982–95 and 1996–2008. For this surface analysis, we use only data from the European subsection of the global domain shown in Fig. 1b. We also exclude grid boxes with less than 90% land because the associated freely evaporating water fraction has a strong effect on gridbox mean surface temperature.

e. *JULES*

JULES (Best et al. 2011; Clark et al. 2011) is the land surface scheme used in HadGEM3-A, which can also be run offline from the GCM. JULES calculates fluxes of energy, water, carbon, and momentum between the land and the atmosphere in response to near-surface boundary conditions of air temperature, humidity, pressure and wind speed, incident short- and longwave radiation, and rain and snow fluxes. The model uses a tiled scheme to represent subgrid heterogeneity in which each grid box is divided into nine tiles of different surface types (five vegetation and four nonvegetation) with a separate surface energy budget calculated for each tile. A static map of tile fractions based on IGBP land-cover classes is used and seasonal variation in vegetation is achieved by prescribing a MODIS-derived climatology of LAI for each vegetation tile.

All tiles in a grid box share the same subsurface soil moisture reservoir, which is split into four soil layers with thicknesses of 0.1, 0.25, 0.65, and 2.0 m, giving a total soil depth of 3 m. Vegetation tiles have roots in all soil layers but include an exponential decay in root density with depth at a rate specific to the vegetation

type. The associated fraction of root mass in each layer affects the total transpiration rate for a vegetation tile and determines what fraction of the transpired water is extracted from each layer. Bare soil evaporation occurs only from the uppermost soil layer using a surface conductance that increases quadratically with soil moisture content. The moisture content of each layer evolves following Richards's equation, with Darcy's equation describing the fluxes between layers. Hydraulic conductivity and matric potential are related to soil moisture content using the relationships of Brooks and Corey (1964), for which spatially varying soil parameters are based on data from the Harmonized World Soil Database.

To drive the offline simulations we use the 3-hourly WFDEI data, which JULES interpolates to the 30-min model time step, using the mean preserving interpolation method of Sheng and Zwiers (1998) for radiation fluxes. Simulations are run from 1979 to 2012, but only output coinciding with the MODIS period of 2000–12 is used in the analysis. Unlike with the HadGEM3-A simulations, it is not necessary to exclude coastal grid boxes from the JULES analysis because both the WFDEI data and the model outputs represent only the land portion of the grid box regardless of the land fraction.

3. Methods

a. Composite LST diagnostics

In this study, we use the dry spell LST-based diagnostics introduced by Folwell et al. (2016) and Gallego-Elvira et al. (2016), which quantify the time-evolving difference between the surface and air temperatures under clear-sky conditions. We focus on the temperature difference rather than the LST alone, as midlatitude LST can be strongly driven by synoptic variations in air temperature. Moreover, the difference in temperatures is proportional to the sensible heat flux, which, under clear-sky conditions, is inversely related to the evaporative fraction. This relationship can be weaker for individual dry spells where evaporation is moderated by atmospheric demand rather than reduced by surface conductance, for example, through moistening of near-surface air as the soil dries. In that situation, the response of sensible heat flux seen in temperature anomalies may not reflect the weaker response of evaporation. But by averaging over many thousands of dry spell events, the dominant mean evaporation response to declining soil moisture can be seen reliably via temperature anomalies.

We define a dry spell as a period of at least 10 days with less than 0.5 mm precipitation each day. Observed dry spells are calculated using WFDEI precipitation from which 3-hourly values are accumulated to daily values from 0000 to 0000 UTC. For the rest of this analysis, we consider only dry spells that begin in the spring and summer months of March–August.

This definition yields a catalog of dry spell events that is used to sample from the 0.5° MODIS LST anomalies. Concurrent dry spells in different grid boxes are treated as separate events. In an individual dry spell, observations are typically too infrequent to reliably infer a dry-down temperature signal above the observation noise. To account for this, we composite the surface and air temperature anomaly data over multiple events as a function of dry spell day. This yields a temperature difference diagnostic (TD) for each dry spell composite day j:

$$\mathrm{TD}_j = \frac{\sum_{i=1}^{n} w_{ij}[(T_{s,ij} - T^c_{s,ij}) - (T_{a,ij} - T^c_{a,ij})]}{\sum_{i=1}^{n} w_{ij}},$$

where n is the total number of dry spells in the composite, T_a and T_s are the air and land surface temperatures, T^c_a and T^c_s are their respective climatology values, and w_{ij} are averaging weights. Each weight w_{ij} is the number of MODIS 1-km pixels used to calculate a 0.5° mean $T_{s,ij}$. This emphasizes days with more complete LST coverage, which are both more representative of the full 0.5° gridbox value and indicative of high levels of insolation. Both land surface and air temperature daily climatologies are calculated using all clear-sky days (i.e., dry spell and nondry spell) from 2000 to 2012 and using the MODIS-derived clear-sky weights w_{ij} before being smoothed using a Gaussian filter. The MODIS weights must be applied to air temperatures as well as LST to avoid biases in TD owing to inconsistent sampling between variables. The same methodology (including the weighting) is used to compute TD from JULES output, but in this case the LST is the full gridbox mean rather than the LST on the gridbox fraction observed by MODIS.

Dry spell events can be grouped by various conditions, such as time and location, prior to compositing to help characterize different surface behaviors. In this study, we stratify events by mixtures of region (based on Köppen–Geiger class, see Fig. 1), season, and antecedent rainfall amount. When doing this, a balance must be achieved between separating the system into finer categories and having enough events in each category to establish the composite signal reliably. For this reason we stratify the JULES analysis by region and three antecedent precipitation categories and the HadGEM3-A analysis by region and two seasons.

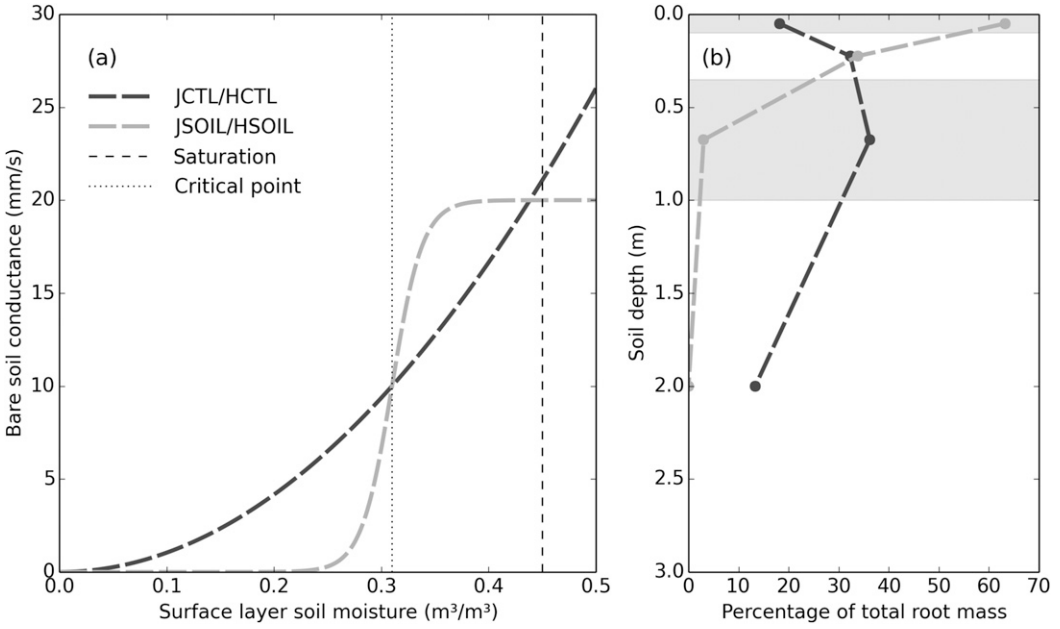

FIG. 2. JULES process revisions: (a) bare soil conductance to evaporation and (b) root mass distribution within the four soil layers.

We also calculate the RWR as the gradient of a linear regression to days 2–11 of a TD composite. Day 1 is excluded from the RWR calculation because it is often strongly influenced by wet canopy evaporation from rainfall on the previous day, a process that is not under consideration here. To assess how TD and RWR vary with initial soil moisture, we stratify dry spells by antecedent rainfall amount before compositing. Rainfall accumulated over 30 days prior to the first dry spell day is used as a simple, observable proxy for root-zone soil moisture availability at the onset of the dry spell. Gallego-Elvira et al. (2016) found that RWR as a function of antecedent rainfall was qualitatively insensitive to choices of accumulation period from 10 to 90 days.

b. *Revised JULES subsurface processes*

One of the aims of this study is to examine the response of TD and RWR to changes in the JULES land surface description both offline and in HadGEM3-A. For these to be useful diagnostics of modeled soil moisture limitation, they should reflect evaporation differences arising from model configuration. Van den Hoof et al. (2013) describe how the soil moisture constraint on evaporation in JULES is likely to be too weak or not to occur often enough, possibly through too much bare soil evaporation (Ebs). This process affects the short-term behavior of evaporation and LST during dry spells and therefore should affect model TD. One way to limit Ebs might be to reduce the thickness of the top soil layer to reduce the reservoir size, but, while this is possible in offline simulations, it is known to be numerically unstable in the MetUM (Van den Hoof et al. 2013). Instead, we modify the surface conductance for bare soil evaporation g_{soil} from

$$g_{\text{soil}} = \frac{1}{100}\left(\frac{\theta_1}{\theta_c}\right)^2$$

to a logistic function

$$g_{\text{soil}} = \frac{1}{50\{1+\exp[-70(\theta_1-\theta_c)]\}},$$

where θ_1 ($\text{m}^3\,\text{m}^{-3}$) is the unfrozen soil moisture content of the upper soil layer and θ_c ($\text{m}^3\,\text{m}^{-3}$), the critical point for that layer, is the soil moisture content at which the matrix potential is -33 kPa. These functions are shown in Fig. 2a. Both functions are constrained to give a conductance of $10\,\text{mm}\,\text{s}^{-1}$ when soil moisture is at the critical point, but the revised function declines more rapidly through this point as the surface dries and offers a greater conductance for high soil moisture contents. A similar function is used in the Community Land Model (Swenson and Lawrence 2014), although it is applied in the surface energy partition in a different way to how g_{soil} is used in JULES.

Because bare soil evaporation and transpiration in JULES are drawn from a common reservoir, any reduction in Ebs is typically compensated by an increase in

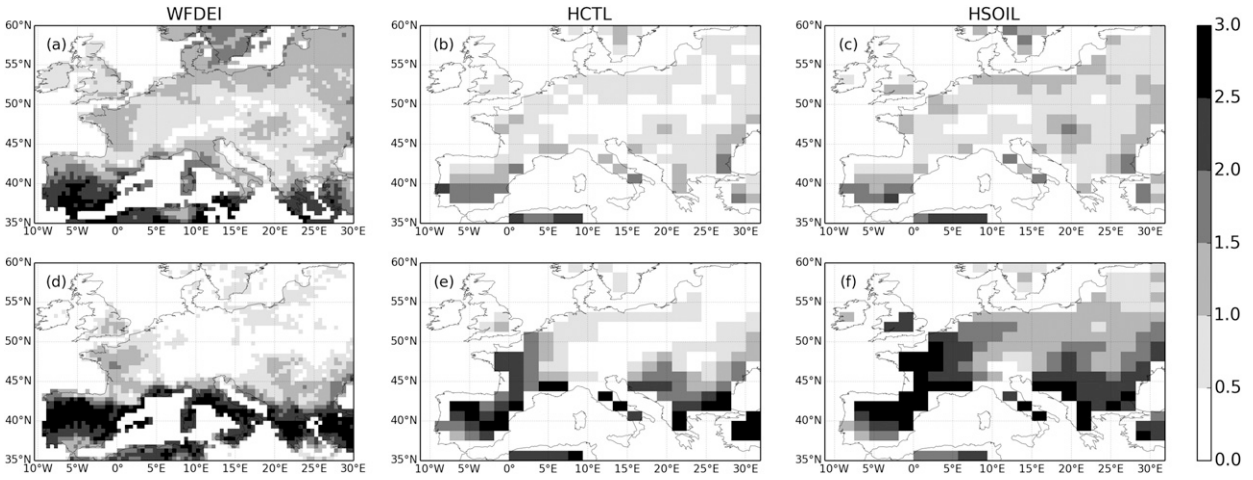

FIG. 3. The observed and HadGEM3-A-simulated number of dry spell events per year in (top) MAM and (bottom) JJA for (a),(d) WFDEI; (b),(e) HCTL; and (c),(f) HSOIL.

evaporation (ET), resulting in little change in the gridbox surface energy budget and LST. To achieve a response in TD it is necessary to also limit transpiration, which we do by reducing the e-folding depth d_r for the root density of grass tiles. The default value of 0.5 m is reduced to 0.1 m, which increases the root mass in the top 0.35 m (the top two soil layers) from 50% to 97% (Fig. 2b). We expect this revised model to show more rapid changes in surface fluxes during the initial days of soil dry down that are of interest here. This root depth value is chosen to improve the fit between modeled and observed TD over the first 15 dry spell days in CWE (not shown). In the MED region, however, a better fit is achieved with the default value. We do not formally optimize this parameter value because we are only assessing how changes to the model surface hydrology can affect TD, and not all of the error in model TD will be due to the processes that we have highlighted.

In this paper, we refer to JULES and HadGEM3-A simulations using the standard surface scheme as JCTL and HCTL respectively, and those simulations using the revised scheme as JSOIL and HSOIL.

c. HadGEM3-A clear-sky sampling

When calculating the TD diagnostic from offline JULES simulations under the historical WFDEI forcing, the clear-sky sampling bias present in the MODIS-based TD is easily replicated by sampling and weighting using the MODIS 1-km pixel availability, as described in section 3a. When creating TD from GCM output, it is necessary to replicate the observed sampling bias, but clear-sky days must be determined from model diagnostics. To achieve this we use simulated surface incident shortwave radiation to exclude days when

$$\frac{S_d}{S_d^{\mathrm{cs}}} < 0.9,$$

where S_d is the gridbox total surface incident shortwave radiation and S_d^{cs} is the surface incident shortwave radiation under clear sky, and each is interpolated to 1030 LT from the 3-hourly HadGEM3-A output. Some effects of this screening process are described in section 4b.

4. Dry spells

a. WFDEI dry spells

Applying the dry spell definition to the 2000–12 WFDEI precipitation data yields a catalog of 81 787 dry spell events that begin in March–August. All parts of the domain contribute to this sample and, while there are similar numbers of events in each calendar month, there is also considerable seasonal variation in their spatial distribution. Figures 3a and 3d show spatial distributions of the mean number of events each year for March–May (MAM; 40 842 events) and June–August (JJA; 35 806 events), respectively. In western Europe and around the Mediterranean, most events occur during the summer. In eastern Europe, where the summers are typically wet, almost all the dry spells occur in spring. The net result for the CWE region is relatively even spatial coverage during spring. Figure 3 also emphasizes the rarity of dry spells at the gridbox level (one or two per year) according to our rather strict definition.

The number and distribution of dry spell events depend on the quality of daily WFDEI precipitation amounts. Folwell et al. (2016) compared the occurrence of WFDEI wet days under this definition with the

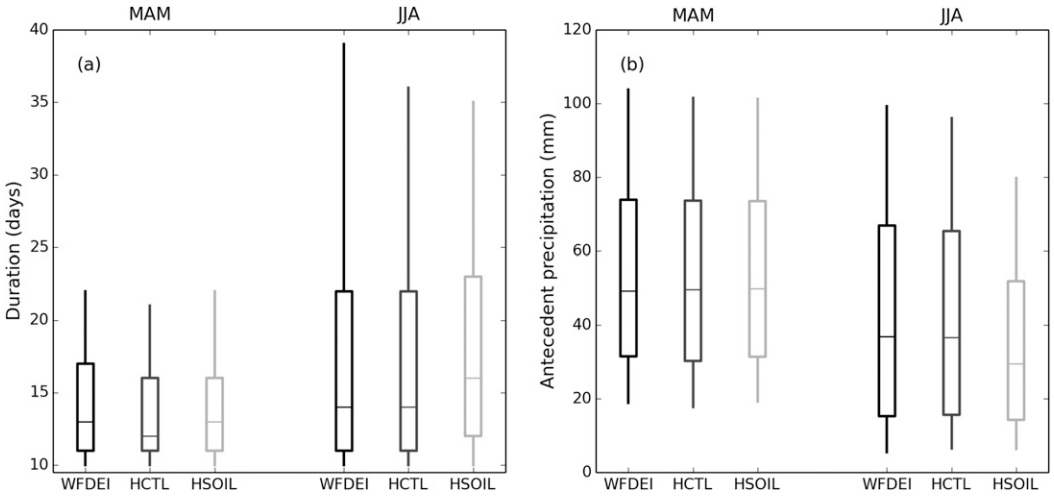

FIG. 4. Summaries of the dry spell event (a) durations and (b) 30-day antecedent precipitation totals, for CWE and MED combined for the seasons MAM and JJA. The boxes span the 25th–75th percentiles, with the median marked by a horizontal line, and the whiskers extend to the 10th and 90th percentiles. HadGEM3-A events are only for grid boxes with more than 90% land.

E-OBS daily gridded station data from the European Climate Assessment and Dataset (Haylock et al. 2008). They found that less than 1% of WFDEI dry spell days were classed as wet days in E-OBS. Contamination of the dry-down temperature signal by rain events not captured in WFDEI, therefore, will affect only a very small percentage of dry spells. This does not exclude situations where we may be rejecting potentially useful events, for example, dry spells that are terminated earlier than necessary or dry spells that are split into parts shorter than 10 days.

b. *HadGEM3-A dry spells*

We applied the same dry spell definition used for WFDEI to the 27 years of HadGEM3-A daily precipitation from the HCTL and HSOIL simulations. In the control simulation this produced a catalog of 10 239 dry spells beginning in March–August, which corresponds to 1.9 dry spells per grid box per year, compared with 2.2 per grid box per year in WFDEI. Figures 3b and 3e show the spatial distribution of these events in MAM (4110 events) and JJA (6129 events). In both CWE and MED, the HadGEM3-A dry spell occurrence rates are slightly too low in spring and slightly too high in summer. These compensate for each other such that HadGEM3-A reproduces the observed March–August occurrence rates in CWE and MED of 1.5 and 3.8 per grid box per year, respectively. Overall, and given the difference in gridbox sizes, HadGEM3-A reproduces very well the observed spatial distribution and seasonal variation in the number of dry spells. The HSOIL simulation produced 14 395 March–August dry spells, corresponding to an increased occurrence rate of 2.8 per grid box per year. The spatial distribution of these events (Figs. 3c,f) shows that this increase in the overall rate comes almost exclusively from CWE. In this region, this corresponds to an improvement in spring and a worsening in summer when dry spells are generated at approximately 3 times the observed rate of 0.55 per grid box per year.

The distributions of the associated observed and modeled dry spell durations are summarized as box plots in Fig. 4a. Qualitatively, HCTL captures the longer dry spells during the summer, but the modeled dry spells are typically too short in spring. For example, the WFDEI and HCTL median durations are both 14 days in JJA and 13 and 12 days, respectively, in MAM. Overall, too few dry spell days are simulated in HCTL because of the combined underprediction in both the occurrence rate and duration of spring dry spells. In the HSOIL simulation, the median durations increase to 13 and 16 days in MAM and JJA, respectively. When considered with the dry spell occurrence rates, this corresponds to a small increase in the number of dry spell days in spring and a substantial increase in the number in summer.

The observed and modeled distributions of 30-day antecedent precipitation are summarized as box plots in Fig. 4b for MAM and JJA. The distribution of antecedent states sampled from the HCTL dry spells is similar to those from the WFDEI-derived dry spells in both seasons. This agreement reflects the combination of a good monthly rainfall climatology for Europe in HadGEM3-A and a good subseasonal distribution of dry periods shown in Figs. 3 and 4. As with the dry spell occurrence rates and durations, the HSOIL antecedent

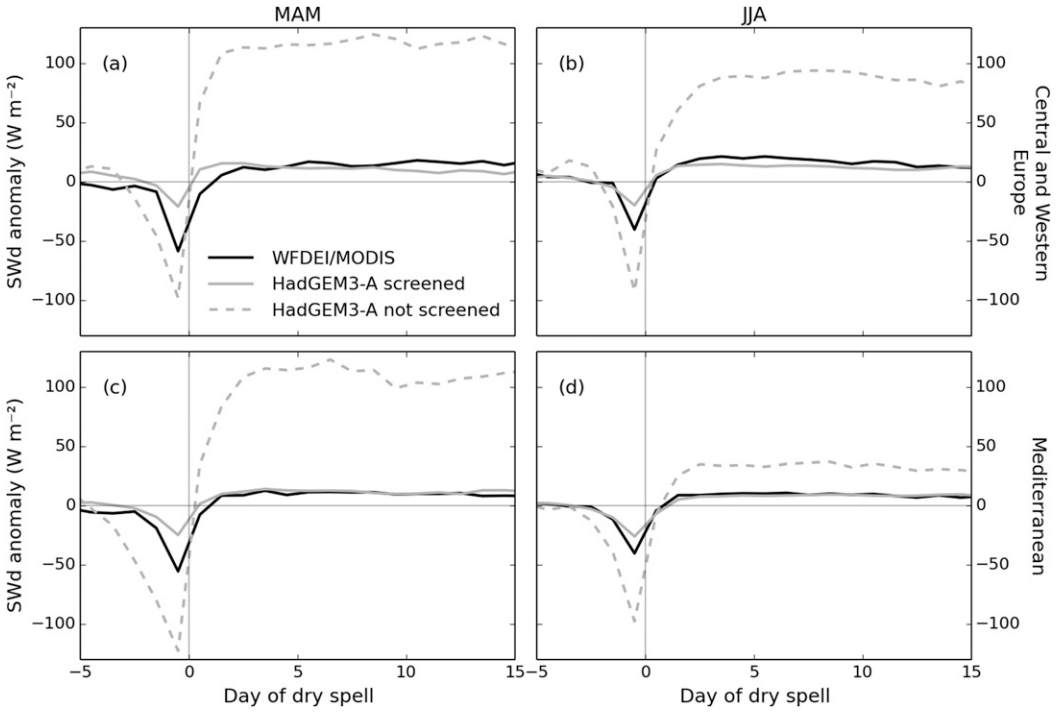

FIG. 5. Observed and HadGEM3-A composite surface incident shortwave radiation anomaly prior to and during dry spells stratified by season and region. Full lines show composites calculated over clear-sky days only, where MODIS availability is used for WFDEI and the method described in section 4b is used for HadGEM3-A. The dashed lines show HadGEM3-A composites without clear-sky screening.

precipitation shows relatively little change from the default HadGEM3-A in spring. However, in summer there is a strong reduction in antecedent precipitation amounts, particularly in the high end of the distribution and despite an increased proportion of events occurring in central and western Europe. This conditional reduction is associated with a 39% decrease in mean summer precipitation over European land in HSOIL compared to HCTL.

To replicate the MODIS clear-sky sampling bias in atmospheric model composites, the screening process described in section 3c is applied to the HadGEM3-A outputs. This screening removes 41% and 19% of dry spell days 1–15 in MAM and JJA, respectively, compared with MODIS removal rates of 50% and 42% based on 1-km pixel LST availability. Figure 5 shows the effect this screening has on HadGEM3-A composite S_d anomalies compared with the equivalent WFDEI-based composites. When the clear-sky screening is included, the observed and simulated dry spell composite S_d anomalies are typically $+10\,\mathrm{W\,m^{-2}}$ in each region and season. On the other hand, when the screening is omitted, the composite radiation anomalies reach around $100\,\mathrm{W\,m^{-2}}$. This is because cloudy, low-radiation days occur more frequently outside of dry spells, increasing the contrast in S_d between dry spell days and the baseline climatology. The effect of the screening is weakest in the MED region during the summer, when clear-sky days are common. Without this screening of HadGEM3-A output, surface and air temperature anomalies would be artificially large. Finally, an important point from Fig. 5 is that composite S_d anomalies are stable through the dry spells, so TD trends over 10 or more days are unlikely to be forced by radiation tendencies.

5. JULES results

a. Composite temperature diagnostics

We first compare how TD evolves in the offline JULES simulations compared to the observations. Dry spell composites of TD and modeled evaporation are shown in Figs. 6 and 7 for the CWE and MED regions, respectively. Folwell et al. (2016) showed that TD can be used to distinguish information about the surface evaporation regime by stratifying dry spell events by antecedent precipitation. In Fig. 6, dry spell events are stratified by terciles of 30-day antecedent precipitation, where the tercile bounds at 31 and 60 mm are calculated

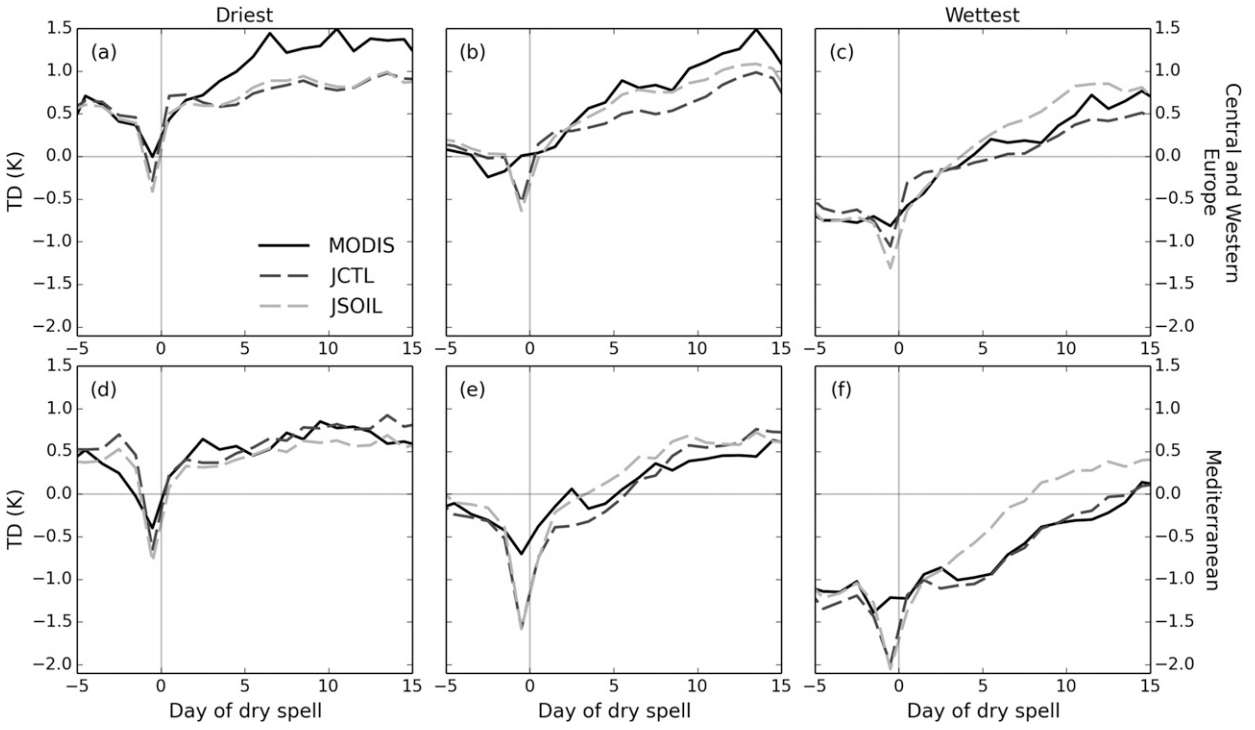

FIG. 6. Dry spell TD in which March–August events are stratified by 30-day antecedent precipitation into (a),(d) dry; (b),(e) intermediate; and (c),(f) wet events, for (top) CWE and (bottom) MED. The precipitation terciles are defined using all March–August events in the European domain.

from all March–August events on the whole European domain. Despite differences in CWE and MED antecedent precipitation described in section 4a, both regions are represented by at least 6600 events in each tercile.

For the driest events in central and western Europe (Fig. 6a), a strong observed RWR of 0.09 ± 0.01 K day^{-1} (mean ± standard error of the mean), suggesting increasing water limitation, contrasts with a weak JULES warming rate in JCTL of 0.02 ± 0.01 K day^{-1} and no change in mean evaporation throughout the dry spell (Fig. 7a). This contrast is even greater over the first few composite days (days 2–6), when the observed and JCTL warming rates are 0.13 ± 0.01 and 0.0 ± 0.03 K day^{-1}, respectively. Unlike the observed behavior for these driest cases, the implication is that often JULES is able to sustain evaporation at rates that do not deplete the soil moisture reservoir enough to reduce the evaporative fraction and induce surface warming. However, this does not allow us to determine whether the required decline in JULES evaporation should come from Ebs or transpiration. For the intermediate and wettest CWE dry spells, a weak decline in JULES composite ET from declining Ebs (Figs. 7b,c) is enough to induce slightly greater RWRs (0.04 ± 0.004 and 0.06 ± 0.005 K day^{-1} in Figs. 6b and 6c, respectively), but this is still weaker than observed for these cases (0.09 ± 0.01 and 0.08 ± 0.01 K day^{-1}).

The weak JULES TD response in CWE across antecedent precipitation conditions (Figs. 6a–c) arises mainly from dry spell events in spring (not shown). While the antecedent precipitation amount controls the mean dry spell evaporation, JULES commonly responds in the short term (10 days) as if there is plenty of water available. However, the observations indicate that water limitation should occur for more of these spring cases. Such behavior could create a tendency for too much spring evaporation leading to summers with overly dry soil moisture conditions and heat waves that are too frequent or too intense.

In the Mediterranean, JULES compares well with the observed TD across the antecedent rainfall strata (Figs. 6d–f). For the wettest events, the initial TD reduction is large and the JULES RWR of 0.10 ± 0.01 K day^{-1} is slightly stronger than the observed rate of 0.08 ± 0.01 K day^{-1}. The observed warming rate over days 2–6 is approximately zero (−0.01 ± 0.02 K day^{-1}) and this is reproduced by JULES (+0.02 ± 0.02 K day^{-1}). An interpretation of this observed pattern of constant TD anomaly followed by rapid warming is of a few days of radiation-limited evaporation followed by a transition into a water-limited regime. For JULES, this is reflected in the evaporation composites shown in Fig. 7f, where

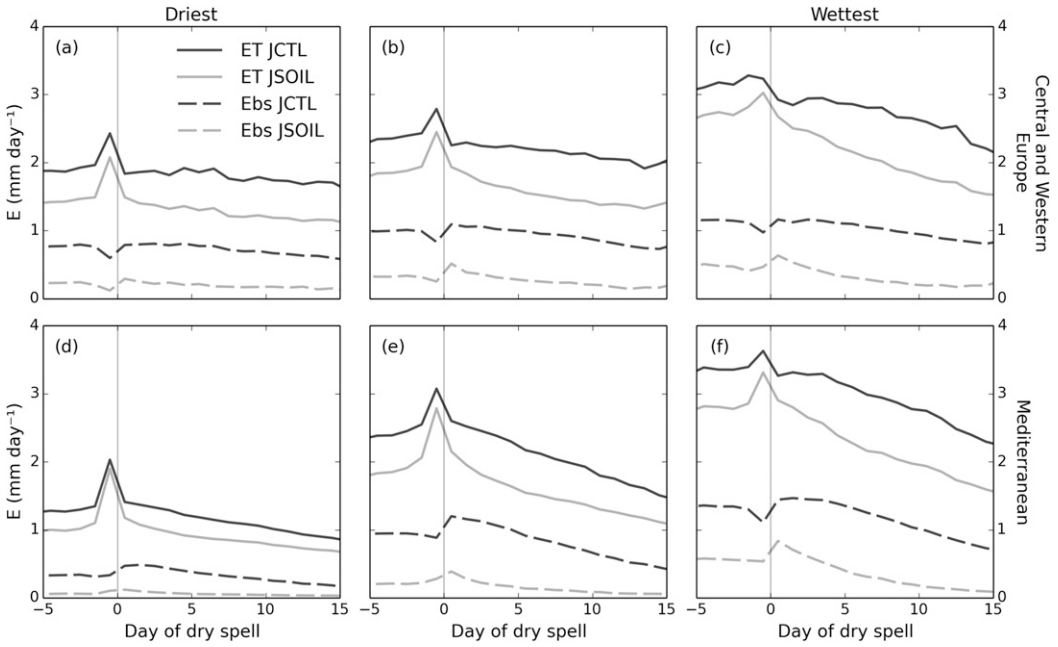

FIG. 7. Dry spell composites of JULES daily mean ET and bare soil evaporation (Ebs; including the contribution from vegetation tiles). Sampling and stratification as per Fig. 6.

total evaporation remains very weak over the first few dry spell days before declining.

For intermediate MED events (Fig. 6e), the observed warming rate ($0.07 \pm 0.01\,\mathrm{K\,day^{-1}}$) is similar to that for the wettest events, whereas the JULES warming rate ($0.12 \pm 0.01\,\mathrm{K\,day^{-1}}$) is much stronger. This indicates that the observed boundary between soil moisture–limited and radiation-limited drying occurs at greater antecedent rainfall amounts than is exhibited by JULES. Once the driest MED events are reached (Fig. 6d) the JULES warming rate ($0.06 \pm 0.01\,\mathrm{K\,day^{-1}}$) has decreased notably from that simulated following wetter conditions and is comparable with the observed rate ($0.04 \pm 0.01\,\mathrm{K\,day^{-1}}$). Note that for these driest events the observed TD returns to a prerainfall event value within a couple of days, indicating that this composite contains occurrences of isolated rainfall within longer dry periods. But even for these particularly dry events JULES composite ET and bare soil evaporation declines only slowly, with mean ET reducing from 1.4 to $1.0\,\mathrm{mm\,day^{-1}}$ over the first 10 days.

A common behavior among the dry spell composites in Fig. 7 is the relatively slow decline in JULES bare soil evaporation during the initial 15 dry spell days. The *e*-folding decay time scales for evapotranspiration fitted against days 1–15 of Fig. 7 are 128, 90, and 57 days for CWE and 30, 27, and 40 days for MED. Of the main evaporation components, the *e*-folding decay time scales for bare soil evaporation are 47, 35, and 38 days for CWE and 14, 15, and 21 days for MED, but for transpiration the declines are too weak for the decay times to be determined reliably.

These decay rates for CWE bare soil evaporation are longer than those of 5 days (sand), 30 days (loam), and 10 days (clay loam) estimated by Brutsaert (2014) from site-level observations. Those JULES rates may also be underestimated because they are fitted against composite absolute evaporation, which is likely to contain the sampling artifact of longer dry spells occurring at climatologically drier locations. An overestimation of bare soil evaporation in JULES was also noted by Van den Hoof et al. (2013) when comparing JULES with FluxNet sites in central Europe. They found that the observed mean Ebs/ET ratios of 0.1–0.2 were lower than the JULES-simulated ratio of 0.34, which is reproduced here with a value of 0.32. While this JULES ratio is typical for land surface models (Dirmeyer et al. 2006), the dry-down Ebs decay is notably slower than that simulated by models with much thinner topsoil layers (e.g., Lawrence et al. 2007).

b. *Response to revised subsurface processes*

The effects of the JULES subsurface changes described in section 3b on TD and ET are also shown in Figs. 6 and 7. In central and western Europe, RWRs increase for the intermediate and wettest terciles (to 0.07 ± 0.01 and $0.13 \pm 0.01\,\mathrm{K\,day^{-1}}$, respectively), with these responses occurring in both spring and summer.

This indicates that, despite the weak TD and evaporation responses in the default set up, JULES has the capacity to strengthen its response even for the wettest group of cases. For the driest cases there is no significant change in the weak warming rate (Fig. 6a), despite a reduction in the composite ET of around 0.5 mm day^{-1} (Fig. 7a). The implication here is that while these JULES modifications reduce mean ET and increase mean temperatures, they do not increase the number of events that experience a decline in ET through the 10 days of dry down. To generate a greater RWR going from JCTL to JSOIL, it would be necessary to increase the number of events with strong ET decline.

The JSOIL TD and ET responses in the MED region are similar to CWE, but the implications relative to the observations are different. The warming rate for the wettest cases increases to 0.15 ± 0.01 K day^{-1} and for the intermediate cases decreases to 0.10 ± 0.01 K day^{-1}, which indicates that the occurrence of soil moisture–limited regimes extends to greater antecedent rainfall amounts. In Fig. 7f, the effect of the g_{soil} change is apparent in the composite means of ET and Ebs, where a uniform evaporation rate in the first few dry spell days is replaced by an immediate decline. In this region, however, the JCTL response for the wettest cases was good overall, so this model revision is inappropriately strong, having an effect on even the highest decile of antecedent precipitation cases during spring and summer. Warming rates decrease slightly for the driest cases to 0.04 ± 0.004 K day^{-1}, but neither this nor the original value is statistically different from the observed value.

These results show that it is possible to change the modeled TD behavior using relatively simple modifications to evaporation processes. The TD responses to the model changes are also qualitatively as expected, and in ways that supports the Folwell et al. (2016) interpretation of observed TD in terms of broad hydrological regimes. Recall that in this offline analysis all the TD diagnostics are calculated relative to the same WFDEI air temperature anomalies, so reflect only differences in the LST anomalies. When JULES is run in HadGEM3-A the combination of land–atmosphere coupling and feedback mechanisms could result in the TD responses being weaker or stronger than those seen offline, and these are the focus of the next section.

6. HadGEM3-A results

a. Composite temperature diagnostics

Figure 8 shows observed and HadGEM3-A-modeled dry spell composite TD for the CWE and MED regions shown in Fig. 1. As described in section 4b, there are substantially fewer dry spells in HadGEM3-A than in the observations, and this limited sample size restricts the amount of stratification that can be done before compositing. For this reason the TD composites in Fig. 8 are stratified by region and season only.

In spring, HadGEM3-A TD values drop to approximately −0.5 K on the last wet day before recovering to around 0 K on the first or second dry spell day and increase steadily thereafter (Figs. 8a,c). The HadGEM3-A relative warming rate in CWE of 0.04 ± 0.01 K day^{-1} is substantially weaker than the observed rate of 0.12 ± 0.02 K day^{-1}, whereas in MED the HadGEM3-A response is similar to the observations (0.10 ± 0.01 and 0.08 ± 0.02 K day^{-1}, respectively). These features are broadly comparable with the JULES results for the intermediate and wettest terciles in Fig. 6. In summer (Figs. 8b,d), the HadGEM3-A TD recovery over the first few days from initially negative values is slower than the observed recovery for both regions. In central and western Europe, the HadGEM3-A RWR is slightly greater than the observed rate (0.08 ± 0.01 and 0.05 ± 0.01 K day^{-1}, respectively), but the TD values are lower on all dry spell days. Similarly, the HadGEM3-A RWR for the Mediterranean is greater than the observed rate (0.09 ± 0.004 vs 0.04 ± 0.01 K day^{-1}). While these RWRs are calculated over days 2–11, each of the HadGEM3-A composites in Fig. 8 show weak increases in TD in the first five dry spell days. This reflects near-constant values of composite evaporative fraction through this period (not shown), with only weak reductions thereafter as the surface dries.

The TD response to soil water availability implied by the stratification into season and region is expressed more directly as a function of antecedent precipitation in Fig. 9. Here RWR is calculated for deciles of 30-day precipitation amount, where the decile bounds are derived using all March–August dry spells in CWE and MED. The observed spring RWRs (Fig. 9a) show lower values for lower precipitation amounts and little change in the responses above 60 mm. The intercepts from the linear regressions are not shown, but the values decrease monotonically with precipitation decile, as was described previously in Folwell et al. (2016). This corresponds to a stronger initial weakening of the surface–air temperature gradient for greater antecedent precipitation amounts, which is consistent with stronger initial ET anomalies in response to greater soil moisture availability.

The weak HadGEM3-A surface response in CWE can be seen as lower than observed RWR values of around 0.05 K day^{-1} for deciles 5–10. Overall, the modeled RWR response weakens slightly with increasing antecedent precipitation whereas the observed response strengthens. These two features are indicative of the

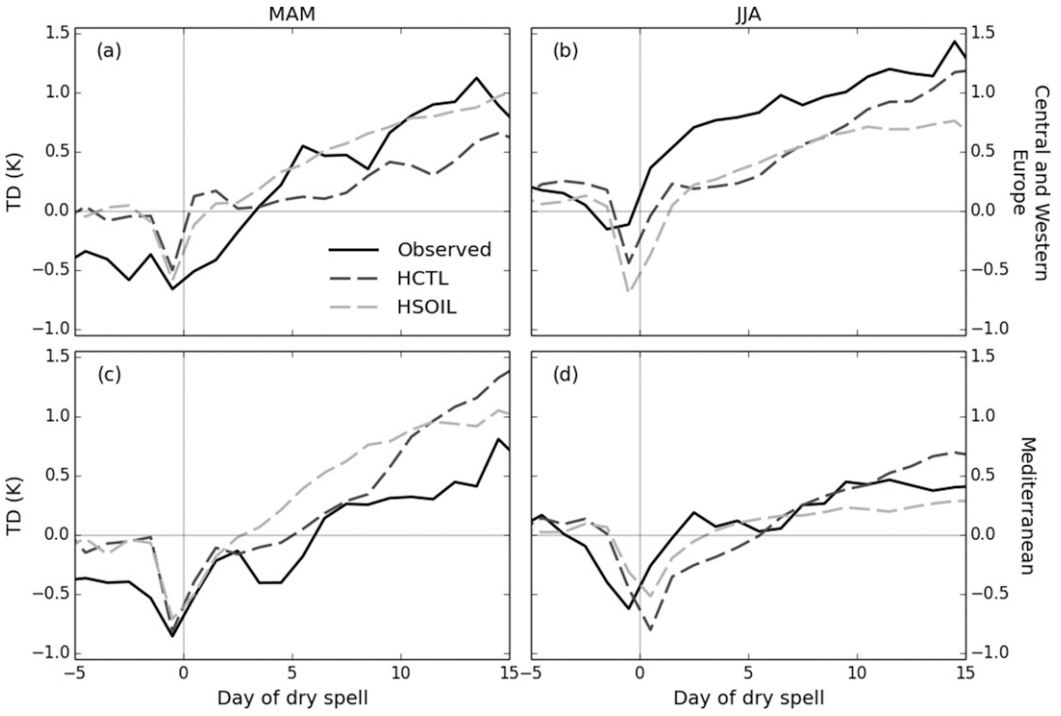

FIG. 8. Observed and HadGEM3-A dry spell composite TD stratified by season (MAM and JJA) and region (CWE and MED).

model exhibiting too little soil moisture limitation on evaporation during the spring, as was suggested by the JULES evaporation in Fig. 7. This is also reflected in composite HadGEM3-A evaporative fraction for CWE (not shown), which remains close to 0.6 through the first 15 dry spell days.

In summer (Fig. 9b), the strong HadGEM3-A response from both regions can also be seen in RWR as a function of antecedent precipitation, where the rates are typically twice as great as those observed for all but the highest and lowest deciles. Overall, the observed response is of moderate RWRs around 0.05 K day^{-1},

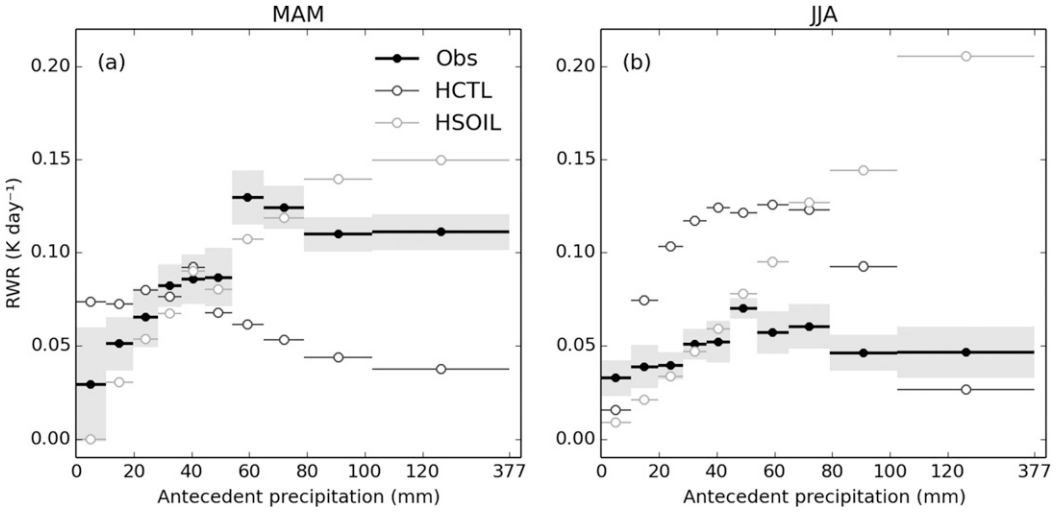

FIG. 9. RWRs for each (a) MAM and (b) JJA dry spell composite as a function of antecedent precipitation decile. Gray bars show standard errors on the RWR values. The decile bounds are calculated from the observed (WFDEI based) March–August events across the whole European domain.

with a weaker dependence on precipitation than is seen in spring. Conversely, HadGEM3-A exhibits a broader range of RWR values (from around 0 to 0.12 K day^{-1}) with a stronger dependence on antecedent precipitation. This relationship is also a demonstration of the idealized bell-shaped curve of RWR as a function of initial soil moisture in Folwell et al. (2016).

b. Response to revised subsurface processes

Results for the HSOIL HadGEM3-A simulation including the bare soil evaporation and root density revisions described in section 3b are shown in Figs. 8 and 9. Unlike with the offline JULES simulations, HSOIL represents the effects of land–atmosphere feedbacks on near-surface air temperature and precipitation that influence the TD analysis. During spring the land surface revisions cause dry spell TD to increase more rapidly in both CWE (0.09 ± 0.004 K day^{-1}) and MED (0.12 ± 0.01 K day^{-1}), as expected. These changes are similar to those seen in the JULES simulations (Fig. 6), so the effects of the surface revisions on TD are not critically weakened by land–atmosphere coupling, that is, dry spell air temperature anomalies do not simply compensate for LST anomalies. Overall, the RWRs as a function of antecedent precipitation (Fig. 9a) are notably improved by the surface revisions in comparison to the observations. Unlike the HCTL simulation, RWRs for HSOIL rise with higher antecedent precipitation, peaking at rates of around 0.15 K day^{-1} for the highest few deciles. This indicates a substantially increased occurrence of soil moisture–limited conditions in HSOIL compared to HCTL. This is reflected in the composite evaporative fraction (not shown), which declines from 0.6 to 0.4 over the first 15 dry spell days. These features suggest that HadGEM3-A is able to exhibit the observed springtime evaporation regimes in Europe.

Results for the European summer are very different. The HadGEM3-A RWR in central and western Europe (0.07 ± 0.004 K day^{-1}) remains relatively unchanged (Fig. 8b) and in the Mediterranean (0.04 ± 0.01 K day^{-1}) is slightly reduced (Fig. 8d), bringing it close to the observed estimate. However, these modest changes in dry spell temperature anomalies are relative to a much drier summer climate. For example, the CWE gridbox dry spell rate (Fig. 3f) increases from 0.88 to 1.6 yr^{-1}, far exceeding the observed rate of 0.55 yr^{-1}. These dry spells are typically associated with longer durations (Fig. 4a) and lower antecedent precipitation amounts (Fig. 4b), as discussed earlier. The overall effect of the surface revisions is that the RWRs increase steadily with antecedent precipitation amount (Fig. 9b), indicating a much greater soil moisture limitation on average. As with the spring, this means that the same high antecedent precipitation forcing results in a stronger TD response in HSOIL than in HCTL.

While this analysis has concentrated on the effect of the surface revisions on dry spell model behavior, the revisions also induce notable changes in seasonal climates. In section 5a, we considered that limiting spring evaporation might increase summer soil moisture and reduce the effects of soil moisture limitation on the surface energy budget. In HSOIL, spring evaporation decreases by 8% in both CWE and MED, and this is associated in summer with a weak 1% increase in top 1-m soil moisture in CWE and a larger 19% increase in MED. Despite these soil moisture increases, summer evaporation decreases by 38% and 31% in CWE and MED, respectively. Any potential shift to wetter summer evaporation regimes by increasing absolute soil moisture is offset by the model revisions, which reduce the availability of soil moisture for evaporation. These revisions introduce a strong cutoff in bare soil conductance with declining soil moisture and effectively restrict water uptake by grasses to the upper two soil levels (0–35 cm depth). In CWE during summer this is compounded by land–atmosphere coupling, which induces a precipitation reduction of 1.0 mm day^{-1} (43%) and an upper 0.35 m soil moisture decrease of 16%, despite the weaker evaporation.

The effects of the HSOIL model revisions on seasonal 2-m air temperatures are shown in Fig. 10. The drier HSOIL evaporation regimes in spring are associated with weak changes in seasonal mean air temperature but stronger increases in the variability of daily mean temperature, particularly in central and western Europe. This reflects the results of Hirschi et al. (2011) and Mueller and Seneviratne (2012), who show that in Europe stronger antecedent precipitation deficits support a wider range in the number of hot summer days per year. In summer, the substantially drier evaporation regimes in HSOIL increase not only the temperature variability (Fig. 10d), but also warm mean temperatures across Europe by 3–5 K (Fig. 10c).

7. Discussion and conclusions

We have shown how the MODIS LST-based diagnostic described by Folwell et al. (2016) and Gallego-Elvira et al. (2016), characterizing the response of the surface energy budget to soil moisture, can be used to examine the behaviors of gridded offline land surface and coupled land–atmosphere models in Europe. This diagnostic is derived using LST from all land-cover classes, so it reflects the combined effects of many land surface processes. For example, it does not separate the responses of bare soil evaporation and transpiration, which Gallego-Elvira et al. (2016) show can differ in this

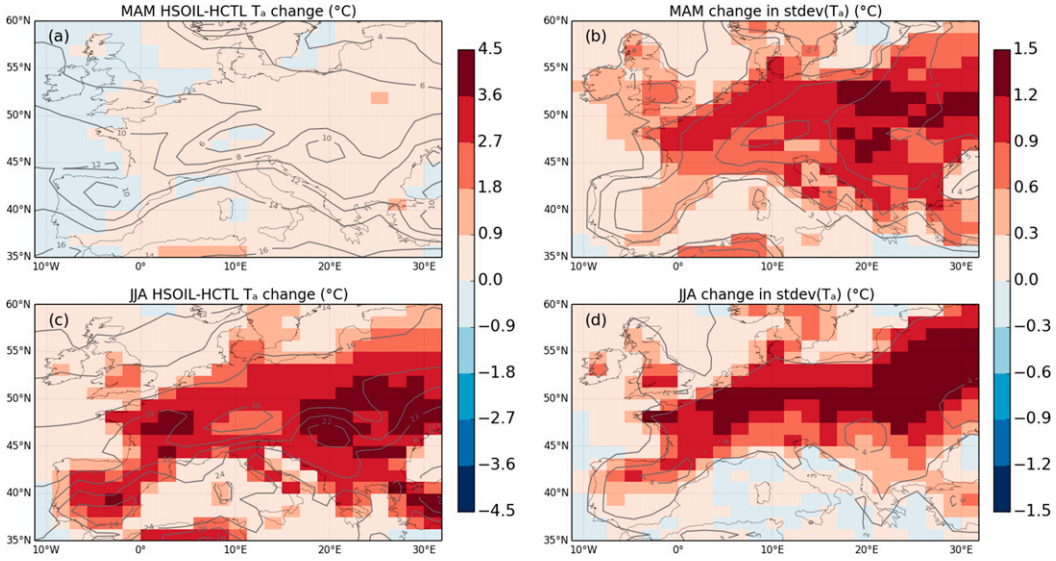

FIG. 10. Changes in 2-m air temperature between HCTL and HSOIL for (top) spring and (bottom) summer: (a),(c) seasonal mean changes and (b),(d) the change in std dev in daily mean temperature. Contours indicate the HCTL base values and colored shading indicates the HSOIL–HCTL changes.

region. Neither does it explicitly separate the direct effect of soil moisture on transpiration from the temperature feedback through stomatal conductance. Nonetheless, it provides an estimate of the overall surface behavior that is useful for evaluating land surface models that do include these processes.

In this study, offline JULES simulations were forced with historical boundary conditions and sampled using the MODIS clear-sky data availability to provide a close comparison with the observed diagnostic. We then showed that HadGEM3-A reproduced the observed properties of European dry spells and that, with appropriate clear-sky day sampling, these could be used to calculate a model equivalent TD from atmospheric GCM simulations. Finally, we assessed the response of HadGEM3-A TD and RWR to revised land subsurface processes, which were identified through offline JULES simulations.

Both JULES and HadGEM3-A reproduced the observed order of dry spell TD values (~1 K) and composite relative warming rates (~0.1 K day^{-1}), although there were notable differences in the regional and seasonal details. The results indicate that HadGEM3-A land surface had too little soil moisture limitation on evaporation in spring and too much in summer, leading to dry spell temperature responses being too weak and too strong, respectively. This interpretation of the European summertime response is consistent with existing results from earlier versions of the MetUM family of models. For example, Stegehuis et al. (2013) and Fischer et al. (2012) showed that regional versions of HadCM3 run for the ENSEMBLES project exhibited low JJA evaporative fraction and overestimated the interannual variability in summer temperature extremes. Similarly, Krueger et al. (2015) found that in HadGEM2-A simulations for CMIP5 hot JJA days had the correct magnitude but were clustered into fewer years than observed. These temperature extremes were also associated too strongly with low relative humidity rather than high surface incident radiation, indicating a tendency toward soil moisture–limited temperature extremes.

The revisions we made to the HadGEM3-A land surface scheme strengthened soil moisture limitation on evaporation during spring, resulting in more realistic RWR responses in central and western Europe. We speculated that reducing evaporation in spring might increase summer soil moisture availability, thereby reducing the range of summer evaporative fraction values. However, while soil moisture increased (despite a significant reduction in summer rainfall), the surface revisions reduced soil moisture availability and evaporation. This possibly invoked similar responses to those identified by Rowell and Jones (2006) for European climate change in an earlier version of MetUM, the HadAM3 with the Providing Regional Climates for Impacts Studies (PRECIS) regional climate model (HadAM3P). They cited spring soil drying as the primary driver of future summer rainfall reduction, followed by land–atmosphere feedbacks. The presence here of a strong coupling between modeled ET and rainfall makes it difficult to determine whether summer RWR behavior would have improved in the absence of that atmospheric response.

While we achieved changes in the modeled RWR behavior through simple revisions to the land surface scheme, this required relatively large parameter changes. The RWR diagnostic is dependent on the model dry spell and climatology behaviors, so it is possible to degrade the model performance in general while improving the performance of anomalies during (relatively infrequent) dry spells. It is also difficult to identify from RWR alone which model process is in error because it is a spatially and temporally aggregated diagnostic and because LST is influenced by properties other than soil moisture availability. For example, we focused here on the modeled soil moisture control, but observed RWR is likely to be affected by irrigation and crop cycles, which were not included in these JULES and HadGEM3-A simulations. Similarly, the specification of land cover, and the associated roughness lengths and albedos, will also contribute to model errors. However, this could be mitigated by calculating observed and modeled RWR separately for different land-cover classes. Given these results, we recommend using RWR as an exploratory tool for characterizing the gross behaviors of land surface and climate models.

Acknowledgments. This study was funded under the Soil Water–Climate Feedbacks in Europe in the Twenty-first Century (SWELTER-21) project (NE/I006729/1) as part of the NERC Changing Water Cycle Programme and through NERC support of the National Centre for Earth Observation. Additional support was funded by the NERC e-stress project (NE/K015990/1). We thank three anonymous reviewers for their insightful comments.

REFERENCES

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2013: Physical constraints for temperature biases in climate models. *Geophys. Res. Lett.*, **40**, 4042–4047, doi:10.1002/grl.50737.

Berg, A., and Coauthors, 2015: Interannual coupling between summertime surface temperature and precipitation over land: Processes and implications for climate change. *J. Climate*, **28**, 1308–1328, doi:10.1175/JCLI-D-14-00324.1.

Best, M. J., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), Model description—Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 595–640, doi:10.5194/gmdd-4-595-2011.

Bisselink, B., E. van Meijgaard, A. J. Dolman, and R. A. M. de Jeu, 2011: Initializing a regional climate model with satellite-derived soil moisture. *J. Geophys. Res.*, **116**, D02121, doi:10.1029/2010JD014534.

Blyth, E., J. Gash, A. Lloyd, M. Pryor, G. P. Weedon, and J. Shuttleworth, 2010: Evaluating the JULES land surface model energy fluxes using FLUXNET data. *J. Hydrometeor.*, **11**, 509–519, doi:10.1175/2009JHM1183.1.

Boberg, F., and J. H. Christensen, 2012: Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nat. Climate Change*, **2**, 433–436, doi:10.1038/nclimate1454.

Boé, J., and L. Terray, 2008: Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change. *Geophys. Res. Lett.*, **35**, L05702, doi:10.1029/2007GL032417.

——, and ——, 2014: Land–sea contrast, soil–atmosphere and cloud–temperature interactions: Interplays and roles in future summer European climate change. *Climate Dyn.*, **42**, 683–699, doi:10.1007/s00382-013-1868-8.

Brooks, R. H., and A. T. Corey, 1964: Hydraulic properties of porous media. Colorado State University Hydrology Paper 3, 27 pp. [Available online at https://dspace.library.colostate.edu/bitstream/handle/10217/61288/HydrologyPapers_n3.pdf?sequence=1.]

Brutsaert, W., 2014: Daily evaporation from drying soil: Universal parameterization with similarity. *Water Resour. Res.*, **50**, 3206–3215, doi:10.1002/2013WR014872.

Christensen, J. H., and F. Boberg, 2012: Temperature dependent climate projection deficiencies in CMIP5 models. *Geophys. Res. Lett.*, **39**, L24705, doi:10.1029/2012GL053650.

——, ——, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophys. Res. Lett.*, **35**, L20709, doi:10.1029/2008GL035694.

Clark, D. B., and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), Model description—Part 2: Carbon fluxes and vegetation. *Geosci. Model Dev.*, **4**, 641–688, doi:10.5194/gmdd-4-641-2011.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, **87**, 1381–1397, doi:10.1175/BAMS-87-10-1381.

——, Y. Jin, B. Singh, and X. Yan, 2013: Trends in land–atmosphere interactions from CMIP5 simulations. *J. Hydrometeor.*, **14**, 829–849, doi:10.1175/JHM-D-12-0107.1.

Ferranti, L., and P. Viterbo, 2006: The European summer of 2003: Sensitivity to soil water initial conditions. *J. Climate*, **19**, 3659–3680, doi:10.1175/JCLI3810.1.

Feudale, L., and J. Shukla, 2007: Role of Mediterranean SST in enhancing the European heat wave of summer 2003. *Geophys. Res. Lett.*, **34**, L03811, doi:10.1029/2006GL027991.

——, and ——, 2011: Influence of sea surface temperature on the European heat wave of 2003 summer. Part II: A modeling study. *Climate Dyn.*, **36**, 1705–1715, doi:10.1007/s00382-010-0789-z.

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:10.1029/2006GL027992.

——, ——, 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.

——, J. Rajczak, and C. Schär, 2012: Changes in European summer temperature variability revisited. *Geophys. Res. Lett.*, **39**, L19702, doi:10.1029/2012GL052730.

Folwell, S., P. P. Harris, and C. M. Taylor, 2016: Large-scale surface responses during European dry spells diagnosed from land

surface temperature. *J. Hydrometeor.*, **17**, 975–993, doi:10.1175/JHM-D-15-0064.1.

Gallego-Elvira, B., C. M. Taylor, P. P. Harris, D. Ghent, K. L. Veal, and S. S. Folwell, 2016: Global observational diagnosis of soil moisture control on the land surface energy balance. *Geophys. Res. Lett.*, **43**, 2623–2631, doi:10.1002/2016GL068178.

Haylock, M. R., N. Hofstra, A. M. G. K. Tank, E. J. Klok, P. D. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:10.1029/2008JD010201.

Hirschi, M., and Coauthors, 2011: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nat. Geosci.*, **4**, 17–21, doi:10.1038/ngeo1032.

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.

Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.*, **116**, D02102, doi:10.1029/2010JD014545.

Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, doi:10.1038/nature09396.

Krueger, O., G. C. Hegerl, and S. F. B. Tett, 2015: Evaluation of mechanisms of hot and cold days in climate models over central Europe. *Environ. Res. Lett.*, **10**, 014002, doi:10.1088/1748-9326/10/1/014002.

Lawrence, D. M., P. E. Thornton, K. W. Oleson, and G. B. Bonan, 2007: The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: Impacts on land–atmosphere interaction. *J. Hydrometeor.*, **8**, 862–880, doi:10.1175/JHM596.1.

Lhotka, O., and J. Kyselý, 2015: Spatial and temporal characteristics of heat waves over central Europe in an ensemble of regional climate model simulations. *Climate Dyn.*, **45**, 2351–2366, doi:10.1007/s00382-015-2475-7.

Lorenz, R., E. L. Davin, and S. I. Seneviratne, 2012: Modeling land–climate coupling in Europe: Impact of land surface representation on climate variability and extremes. *J. Geophys. Res.*, **117**, D20109, doi:10.1029/2012JD017755.

——, ——, D. M. Lawrence, R. Stöckli, and S. I. Seneviratne, 2013: How important is vegetation phenology for European climate and heat waves? *J. Climate*, **26**, 10 077–10 100, doi:10.1175/JCLI-D-13-00040.1.

Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman, 2011: Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, **15**, 453–469, doi:10.5194/hess-15-453-2011.

——, M. J. den Berg, A. J. Teuling, and R. A. M. Jeu, 2012: Soil moisture–temperature coupling: A multiscale observational analysis. *Geophys. Res. Lett.*, **39**, L21707, doi:10.1029/2012GL053703.

——, 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. *Nat. Geosci.*, **7**, 345–349, doi:10.1038/ngeo2141.

Mueller, B., and S. I. Seneviratne, 2012: Hot days induced by precipitation deficits at the global scale. *Proc. Natl. Acad. Sci. USA*, **109**, 12 398–12 403, doi:10.1073/pnas.1204330109.

——, and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.*, **38**, L06402, doi:10.1029/2010GL046230.

——, and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, **17**, 3707–3720, doi:10.5194/hess-17-3707-2013.

Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644, doi:10.5194/hess-11-1633-2007.

Quesada, B., R. Vautard, P. Yiou, M. Hirschi, and S. I. Seneviratne, 2012: Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nat. Climate Change*, **2**, 736–741, doi:10.1038/nclimate1536.

Rowell, D. P., and R. G. Jones, 2006: Causes and uncertainty of future summer drying over Europe. *Climate Dyn.*, **27**, 281–299, doi:10.1007/s00382-006-0125-9.

Schär, C., P. L. Vidale, D. Luthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, doi:10.1038/nature02300.

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

——, and Coauthors, 2013: Impact of soil moisture–climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophys. Res. Lett.*, **40**, 5212–5217, doi:10.1002/grl.50956.

Sheng, J., and F. Zwiers, 1998: An improved scheme for time-dependent boundary conditions in atmospheric general circulation models. *Climate Dyn.*, **14**, 609–613, doi:10.1007/s003820050244.

Stéfanon, M., P. Drobinski, F. D'Andrea, and N. de Noblet-Ducoudré, 2012: Effects of interactive vegetation phenology on the 2003 summer heat waves. *J. Geophys. Res.*, **117**, D24103, doi:10.1029/2012JD018187.

Stegehuis, A. I., R. Vautard, P. Ciais, A. J. Teuling, M. Jung, and P. Yiou, 2013: Summer temperatures in Europe and land heat fluxes in observation-based data and regional climate model simulations. *Climate Dyn.*, **41**, 455–477, doi:10.1007/s00382-012-1559-x.

Swenson, S. C., and D. M. Lawrence, 2014: Assessing a dry surface layer–based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. *J. Geophys. Res. Atmos.*, **119**, 10 299–10 312, doi:10.1002/2014JD022314.

Teuling, A. J., S. I. Seneviratne, C. Williams, and P. A. Troch, 2006: Observed timescales of evapotranspiration response to soil moisture. *Geophys. Res. Lett.*, **33**, L23403, doi:10.1029/2006GL028178.

——, and Coauthors, 2009: A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.*, **36**, L02404, doi:10.1029/2008GL036584.

Trigo, I. F., I. T. Monteiro, F. Olesen, and E. Kabsch, 2008: An assessment of remotely sensed land surface temperature. *J. Geophys. Res.*, **113**, D17108, doi:10.1029/2008JD010035.

Van den Hoof, C., P. L. Vidale, A. Verhoef, and C. Vincke, 2013: Improved evaporative flux partitioning and carbon flux in the land surface model JULES: Impact on the simulation of land surface processes in temperate Europe. *Agric. For. Meteor.*, **181**, 108–124, doi:10.1016/j.agrformet.2013.07.011.

Vautard, R., and Coauthors, 2007: Summertime European heat and drought waves induced by wintertime Mediterranean

rainfall deficit. *Geophys. Res. Lett.*, **34**, L07711, doi:10.1029/2006GL028001.

——, and Coauthors, 2013: The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dyn.*, **41**, 2555–2575, doi:10.1007/s00382-013-1714-z.

Walters, D. N., and Coauthors, 2014: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geosci. Model Dev.*, **7**, 361–386, doi:10.5194/gmd-7-361-2014.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, doi:10.1002/2014WR015638.

Weisheimer, A., F. J. Doblas-Reyes, T. Jung, and T. N. Palmer, 2011: On the predictability of the extreme summer 2003 over Europe. *Geophys. Res. Lett.*, **38**, L05704, doi:10.1029/2010GL046455.

Zampieri, M., F. D'Andrea, R. Vautard, P. Ciais, N. de Noblet-Ducoudré, and P. Yiou, 2009: Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *J. Climate*, **22**, 4747–4758, doi:10.1175/2009JCLI2568.1.