

# Geophysical Research Letters



## RESEARCH LETTER

10.1029/2019GL082962

### Key Points:

- The relative warming rate diagnostic is used to assess soil drying impacts on surface heat fluxes in climate models
- This diagnostic is shown to be a very good indicator of modeled evaporative fraction evolution during dry downs
- Models capture dry spell evapotranspiration dynamics well in arid zones but struggle in continental climate zones

### Supporting Information:

- Supporting Information S1

### Correspondence to:

B. Gallego-Elvira,  
belengallegoelvira@gmail.com

### Citation:

Gallego-Elvira, B., Taylor, C. M., Harris, P. P., & Ghent, D. (2019). Evaluation of regional-scale soil moisture-surface flux dynamics in Earth system models based on satellite observations of land surface temperature. *Geophysical Research Letters*, 46, 5480–5488. <https://doi.org/10.1029/2019GL082962>

Received 22 MAR 2019

Accepted 30 APR 2019

Accepted article online 6 MAY 2019

Published online 29 MAY 2019

## Evaluation of Regional-Scale Soil Moisture-Surface Flux Dynamics in Earth System Models Based on Satellite Observations of Land Surface Temperature

Belen Gallego-Elvira<sup>1,2</sup> , Christopher M. Taylor<sup>1,3</sup> , Phil P. Harris<sup>1,3</sup> , and Darren Ghent<sup>3,4</sup>

<sup>1</sup>Centre for Ecology and Hydrology, Wallingford, UK, <sup>2</sup>Agricultural Engineering Department, Technical University of Cartagena, Cartagena, Spain, <sup>3</sup>National Centre for Earth Observation, Wallingford, UK, <sup>4</sup>Department of Physics and Astronomy, University of Leicester, Leicester, UK

**Abstract** There is a lack of high-quality global observations to evaluate soil drying impacts on surface fluxes in Earth system models (ESMs). Here we use a novel diagnostic based on the observed warming of the land relative to the atmosphere during dry spells (relative warming rate, RWR) to assess ESMs. The ESMs show that RWR is well correlated with changes in the partition of surface energy between sensible and latent heat across dry spells. Therefore, comparisons between observed and simulated RWR reveal where models are unable to capture a realistic soil moisture-heat flux relationship. The results show that in general, models simulate dry spell ET dynamics well in arid zones while decreases in evaporative fraction appear excessive in some models in continental climate zones. Our approach can help guide land model development in aspects that are key in simulating extreme events like heat waves.

**Plain Language Summary** We present a methodology to assess how land evapotranspiration (ET) responds to soil moisture (SM) in climate models across the world. Our method is based on looking at the observed versus modeled evolution of land surface temperature and the overlying air temperature during rain free periods, when SM decreases and limits the amount of land ET. We observed a good relationship between modeled ET and the modeled relative evolution of these temperatures during dry periods. Therefore, our method can be used to evaluate how realistic are modeled SM-ET relationships when soils are drying. Our results show that in general, models capture fairly well the influence of SM on ET in arid climate zones, while they perform less well in continental climate zones. Our method provides a unique new tool to improve aspects of weather and climate models important for simulating phenomena such as heat waves.

## 1. Introduction

Soil moisture (SM) plays a crucial role in water and energy partitioning within the Earth system. In water-limited regions, SM fluctuations strongly control the ratio of latent to sensible heat flux. Dry soils can amplify near-surface air temperature ( $T_a$ ) extremes through feedbacks with evapotranspiration (ET; Miralles et al., 2014; Teuling, 2018) and influence rainfall generation (Koster et al., 2004; Taylor et al., 2012). Considering that large changes in water availability are expected later in this century, SM-atmosphere feedbacks will likely affect  $T_a$  and precipitation projections (Seneviratne et al., 2013). This is particularly important for the occurrence of extreme events that have devastating socioeconomic and environmental impacts (Seneviratne et al., 2012).

The representation of SM feedbacks in Earth system models (ESMs) is limited by the poor depiction of ET in land surface schemes (Gevaert et al., 2018). Simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5) disagree on when, where, and how much ET is limited by SM (Mueller & Seneviratne, 2014). Levine et al. (2016) found that land-atmosphere coupling strength was stronger in models than in satellite observations and Dirmeyer et al. (2018) showed that models tend to overestimate correlations between SM and ET when comparing to flux tower observations. Sippel et al. (2017) showed that most CMIP5 simulations systematically produce too frequent coincidences of high  $T_a$  anomalies with negative ET anomalies in midlatitude regions during the warm season and in several tropical regions year-round. Looking at the evaporative fraction (EF) during heat extremes, Ukkola et al. (2018) also found that ESMs overestimate the strength of land-atmosphere coupling, causing the overamplification of heat extremes in wet regions.

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Evaluation of the relevant modeled processes still suffers from a lack of reliable, global observations of ET at the ESM grid box scale. Several studies have presented observation-based metrics of varying complexity for describing different aspects of the coupling between SM and atmospheric conditions. These metrics include correlations between ET and  $T_a$  (Seneviratne et al., 2006), ET and solar radiation (Teuling et al., 2009), an index of surface flux sensitivity to SM (Dirmeyer, 2011), the dependence of EF on SM (Schwingshackl et al., 2017), and coincidences between anomalies of  $T_a$  and ET (Sippel et al., 2017) and of ET and photo-synthetically active radiation (Zscheischler et al., 2015).

Following the approach of Teuling et al. (2006) in assessing land surface behavior during dry downs (rain-free periods), Gallego-Elvira et al. (2016; hereafter GE16) presented the global diagnostic relative warming rate (RWR; mean increase in land surface temperature [LST] relative to  $T_a$  during dry spells) to evaluate SM effects on land fluxes based on satellite observations and reanalysis data. The novelty of the RWR is that it highlights the direct thermal response of the surface energy balance to SM. Alternative global observation-based approaches either link SM to atmospheric variables (Levine et al., 2016; Mueller & Seneviratne, 2012) or rely on assumptions about the SM-ET relationship to interpret the observed variable (Miralles et al., 2012).

In this study, we assess the potential of the RWR diagnostic for the evaluation of regional-scale SM-surface flux relationships in ESM simulations. We derived global modeled RWRs from 10 ESMs and compare them to observed RWRs computed from 15 years of satellite observations and reanalysis data.

## 2. Theoretical Background

The dynamics of LST and  $T_a$  during dry spells, when negligible rain falls, can be used to characterize the changing partition of energy between sensible and latent heat. When soils dry out sufficiently, ET becomes limited, leading to increases in LST (direct response) and  $T_a$  (indirect response, via enhanced surface warming of the planetary boundary layer). In fact, as shown recently by Panwar et al. (2019), the imprint of evaporative conditions on temperature is stronger at the land surface than in the atmosphere. Thus, the rate of land surface warming relative to the atmosphere (RWR) emphasizes the impact of changes in SM control on the surface energy balance. For theoretical illustration, we presented in GE16 a simple model to simulate LST evolution during dry downs across the drying stages I–III, as in Teuling et al. (2010) and SM and ET regimes, as in Seneviratne et al. (2010). The sensitivity of RWR to initial SM can be described by a bell-shaped curve (supporting information Figure S1). In the wet, energy-limited regime (stage I), ET is unconstrained by SM, and there is no change in LST during the dry spell (for constant atmospheric conditions,  $RWR = 0$ ). In the transitional regime (stage II), ET decreases over time as the soil dries, and RWR thus increases. If the onset of the dry spell occurs when the soil is already relatively dry, the decrease in ET over time will be relatively small. This implies that as soils approach stage-III drying, RWR falls toward 0 again. Observations of how RWR varies with soil wetness thus gives an indication of which regimes a region typically experiences. For example, in an arid (wet) region, RWR will rise (fall) with increasing initial SM.

## 3. Methods and Data

### 3.1. Observational and Model Data

The observed RWRs were derived from satellite and reanalysis data. We use clear-sky daytime LST observations from MODIS (Moderate Resolution Imaging Spectroradiometer, on board of Terra and Aqua satellites) for the period 2000 to 2014 (Collection-5) and aggregated from  $0.01^\circ$  to  $0.5^\circ$ . We used 3-hourly  $T_a$  from European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis (Dee et al., 2011), regrided to a regular  $0.5^\circ$  grid and linearly interpolated to the average satellite scan time of each  $0.5^\circ$  grid box. For both land and air temperatures, we use anomalies (relative to a long-term monthly mean) to minimize the impact of variable spatial sampling in regions with strong temperature gradients and to allow us to combine data from the Terra (late morning) and Aqua (early afternoon) satellites. The anomalies of both (surface and air) temperatures are defined using the same method and sampling. It should be noted that by using temperature anomalies, the RWR effectively filters out slowly varying climatological changes in EF and focuses on anomalous higher-frequency land dynamics.

Satellite precipitation was used to identify rain-free periods. We define a dry spell as a sequence of at least 10 days in which the rainfall of all the three following data sets is below 0.5 mm: Tropical Rainfall Measuring

Mission (TRMM, Huffman et al., 2007), Climate Prediction Center MORPHing (CMORPH, Joyce et al., 2004), and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN, Sorooshian et al., 2008). We considered that 10 days is enough to observe the LST response to soil drying while keeping a sufficient quantity of dry spells to enable the analysis globally. The SM condition at the start of each dry spell was estimated from the rainfall accumulation over the previous 30 days. Further details on observational data processing and filtering can be found in GE16.

We evaluated a subset of ten ESMs running the Atmospheric Model Intercomparison Project experiment. The models were selected based on the availability of 3-hourly data of the variables required for the analysis: LST,  $T_a$ , precipitation, latent and sensible heat fluxes, and surface downward solar radiation (rsds) plus clear-sky rsds (rsdscs) or cloud area fraction. Only six models in the CMIP5 archive provide subdaily LST. The data from the additional four ESMs were provided by partners within the framework of the EU FP7 EMBRACE (Earth system Model Bias Reduction and assessing Abrupt Climate change) project as summarized in Table S1. As with the observations, anomalies in simulated land and air temperature are used. When comparing with observed RWRs, this approach reduces the impact of systematic biases in the models and instead emphasizes the short-term evolution of the SM-ET relationship over a dry spell.

### 3.2. Computation of RWRs and EF Gradients

To calculate observed RWRs, the first step is to compute the composite mean difference between LST and  $T_a$  anomalies for each dry spell day ( $LST_{\text{anom}} - T_{a,\text{anom}}$ , equation (1)).

$$LST_{\text{anom}} - T_{a,\text{anom}} = \frac{\sum_{i=1}^n w_{ij} \left[ (LST_{ij} - LST_{ij}^c) - (T_{a,ij} - T_{a,ij}^c) \right]}{\sum_{i=1}^n w_{ij}}, \quad (1)$$

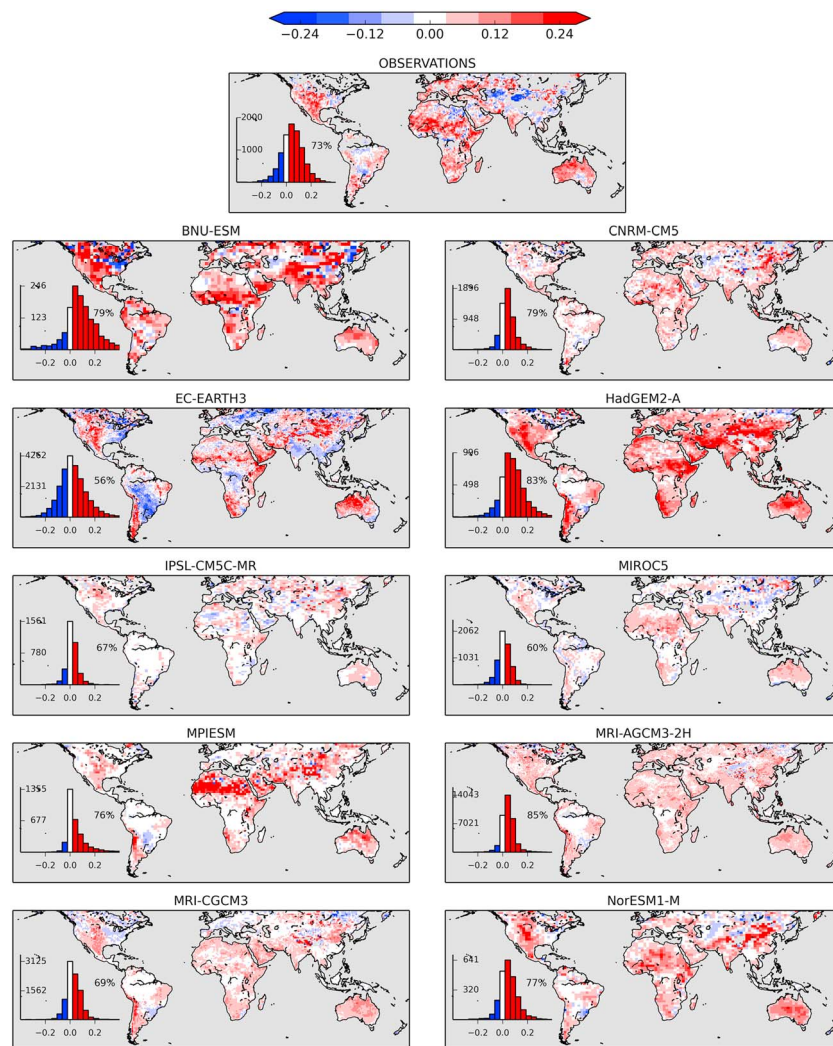
where  $j$  is the dry spell day,  $n$  is the number of dry spell events,  $i$  is a single event with a valid LST observation,  $LST^c$  and  $T_a^c$  are the climatological LST and  $T_a$ , and the weighting factor  $w$  is the number of  $0.01^\circ$  observations used to construct the  $0.5^\circ$  LST. Values of  $T_a$  are only selected on days with a valid  $0.5^\circ$  LST value, so  $LST^c$  and  $T_a^c$  represent clear-sky climatologies. We removed dry spell events with days in which  $T_a$  was below  $10^\circ\text{C}$  at satellite overpass time to avoid artefacts from soil freezing. The second step is to retrieve RWR as the slope of the linear regression of  $LST_{\text{anom}} - T_{a,\text{anom}}$  on days 2 to 10 of the dry spell. The first day of each dry spell was discarded because of lower number of LST observations, timing errors in precipitation, and to exclude days affected by water on the vegetation canopy. Note that during long dry downs, only the 10 first days are considered for the analysis.

To derive ESM RWRs, model 3-hourly output was processed to replicate the clear-sky sampling of LST during dry spells. We considered clear-sky conditions when  $\text{rsds}/\text{rsdscs} > 0.9$  or the cloud area fraction was below 10% at the average satellite overpass local time for MODIS Terra and Aqua combined ( $\sim 1200$  LT). The rest of the computations were analogous to the observations, but they were performed on the native grid of each model. In addition to the variables needed to retrieve RWRs, clear-sky anomalies of modeled fluxes during dry spells were computed to look at the dry-down evolution of the EF. In the same way as for RWRs, we first calculated the composite mean EF anomalies for each dry spell day, and the EF gradient was computed by linear regression of EF on days 2 to 10 of the dry spell.

Finally, in order to assess the sensitivity of RWRs and EF gradients to SM at the onset of a dry down, dry spell events were stratified by observed 30-day antecedent precipitation in decile bins. We studied this sensitivity at regional level using 20 SREX regions defined in Seneviratne et al. (2012), where there were sufficient dry spells and LST observations to enable the analysis. A dominant climatic type was assigned to each SREX region based on the most frequent grid cell class according to the Köppen-Geiger classification (Kottek et al., 2006); equatorial (As, Aw, and Cw); arid cold (Bk) and hot (Bh); temperate humid (Cf) and dry (Cs); and continental hot/warm summer (Dfa and Dfb; hereafter continental).

## 4. Results

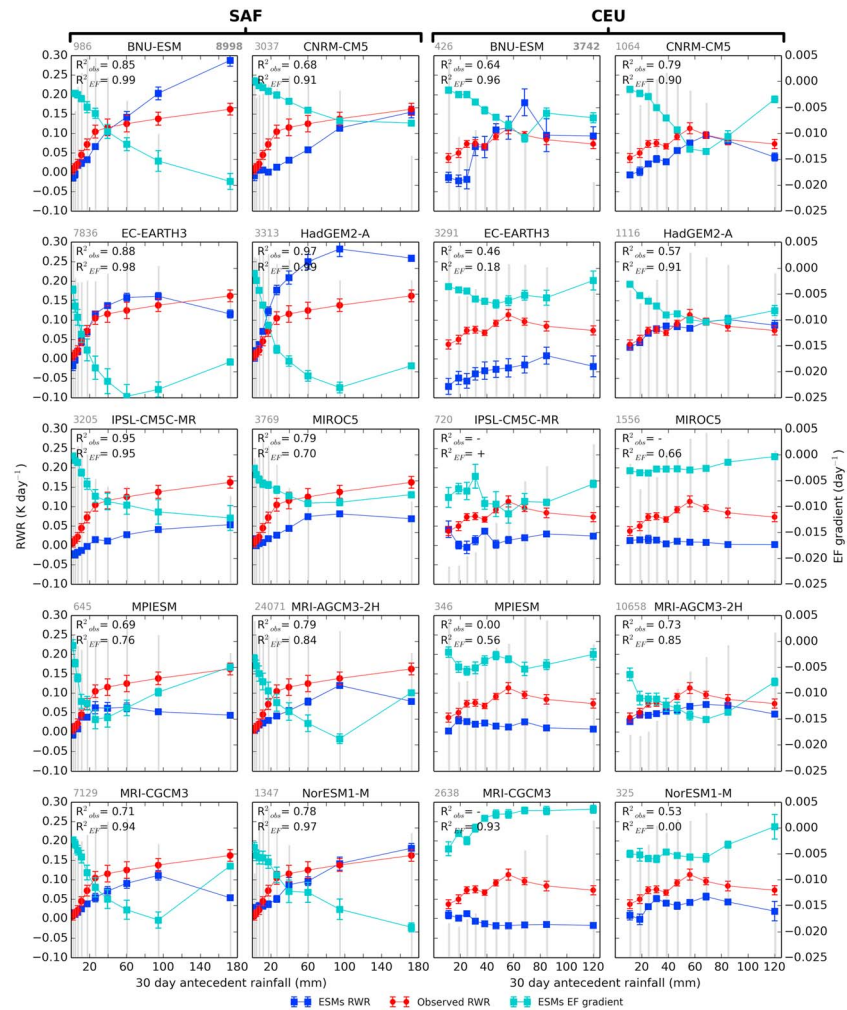
Global observed and modeled RWRs (between  $60^\circ\text{S}$  and  $60^\circ\text{N}$ ) are presented in Figure 1 at  $1^\circ$  resolution for observations and on the native grid for each model. Globally, the area with positive RWR dominates in both observations (73% of the grid boxes) and ESM simulations (from 56% to 85% for EC-EARTH and MRI-



**Figure 1.** Observed and modeled rate of surface warming relative to overlying air during dry spells (relative warming rate [RWR], K/day). Observed RWRs correspond to the 2000–2014 mean at  $1^\circ$  resolution. Modeled RWRs correspond to the mean of AMIP simulations for the periods and model grid resolutions indicated in Table S1. Inset bar plots show the distribution of RWR across all grid cells and are annotated with the percentage of grid cells with positive RWR. Land grid cells with less than one event per year with suitable data for the analysis are shaded gray.

AGCM3-2H, respectively). Positive RWR means that the land surface warms (on average) faster than the atmosphere during 10-day dry downs. Given no clear trend in solar radiation and wind speed over the dry spell (as shown in GE16), positive RWRs can be associated with increases in sensible heat flux. There are some common broadscale features in all models and observations, notably the widespread positive RWR in dry regions such as Australia, southern Africa, and west North America and the dominance of small or even negative RWR in wet and forested regions. Negative RWRs in these areas indicate that there is abundant water in the soil during dry downs for evaporative cooling, and therefore, the air warms faster than the surface. However, in the midlatitudes we can detect several areas of disagreement (in sign and spatial pattern of RWR) among the models and compared to observations: for example, Central Europe (CEU), Eastern North America, and Eastern Asia. In these areas, transitional and wet ET regimes are typical depending on the season (Schwingshackl et al., 2017).

While illustrative of gross properties of the models, a drawback in directly comparing simulated and observed RWR in Figure 1 is that there are a number of factors beyond underlying SM-ET coupling that control RWR. For instance, there is a large range of global mean RWR across the models (e.g., HadGEM2-A and MRI-AGCM3-2H). Large contrasts in mean RWR may reflect different aerodynamic coupling between



**Figure 2.** Observed versus modeled relative warming rate diagnostic (RWR, K day<sup>-1</sup>) and modeled evaporative fraction (EF) gradient (day<sup>-1</sup>) stratified by 30-day observed antecedent precipitation in decile bins for the 10 Earth system models described in Table S1. Error bars correspond to the standard error of the estimated gradient of the linear least squares regression. Plots in columns 1 and 2 correspond to the SREX region Southern Africa (SAF) and columns 3 and 4 to Central Europe (CEU). Gray bars show the distribution of dry spell events produced by the models for each observed antecedent precipitation bin. The maximum number for modeled events per bin is indicated in gray on the top left corner of each plot and for observed events is shown in the top right of the first plot for each region. Correlations between simulated RWR and observations ( $R^2_{obs}$ ) and EF gradient ( $R^2_{EF}$ ) are provided for each model, except when  $R_{obs} < 0$  or  $R_{EF} > 0$ , which is indicated with “-” or “+,” respectively.

models or different approaches to representing subgrid land surface properties. Also, biases in precipitation forcing from an individual ESM may place the land surface in the wrong drying regime. To provide more insight into the modeling of SM-ET coupling, we examine the variation of RWR under different hydrological conditions across a region. Following GE16, we compute RWR as a function of precipitation in the previous 30 days. This proxy of pre-dry spell hydrological status is chosen as, unlike SM, it has a common definition across the models and the observations.

Figure 2 illustrates this analysis for two SREX regions with contrasting hydroclimatic regimes, Southern Africa (SAF) and CEU. The same plot for the remaining SREX regions is available in Figures S2a–S2r. The data from each dry spell and grid cell are grouped in 10 bins defined by the observed deciles of pre-dry spell antecedent precipitation, and these bins are also applied to the models. This allows a comparison between the simulated and observed RWR that is insensitive to monthly rainfall biases in individual models and provides information about the drying stage of the land surface (Seneviratne et al., 2010; Teuling et al., 2010). In SAF, observed RWR is strongly sensitive to antecedent rain, particularly following relatively dry

months (<30 mm/month) and has a maximum value in the 10th decile. This is consistent with dry spell events falling within drying stages II (SM constrains ET) to III (SM is unavailable for ET). All the models exhibit this semiarid behavior to some extent, which is quantified by the correlation between observed and simulated RWR ( $R^2_{\text{obs}}$ ). The relationship between RWR and surface flux dynamics in the models is also illustrated in Figure 2, using the mean gradient of EF over time during dry spells. For each model in SAF, Figure 2 shows that there is a strong negative correlation between RWR and EF gradient. Four of the models (BNU-ESM, CNRM-CM5, IPSL-CM5C-MR, and NorESM1-M) depict the strongest dry spell decrease in EF (most negative value) in the 10th decile, suggesting no influence of stage I drying. These models also have the highest RWR in the 10th decile, consistent with the observations. The relationships between EF gradient and rainfall in the remaining models appear more like inverted bell curves; the EF gradient rises (becomes less negative) under the rainiest conditions. This is indicative of a weakening of the SM control on ET at the wet end and is also reflected in the shape of the RWR curve in these six models. The negative correlation between RWR and EF gradient in each model highlighted here demonstrates the validity of our approach in using RWR as a proxy for dry spell flux dynamics. In supporting information Figure S3 this correlation is quantified ( $R^2_{\text{EF}}$ ) for all models and all regions, with 81% of model regions having values exceeding 0.7.

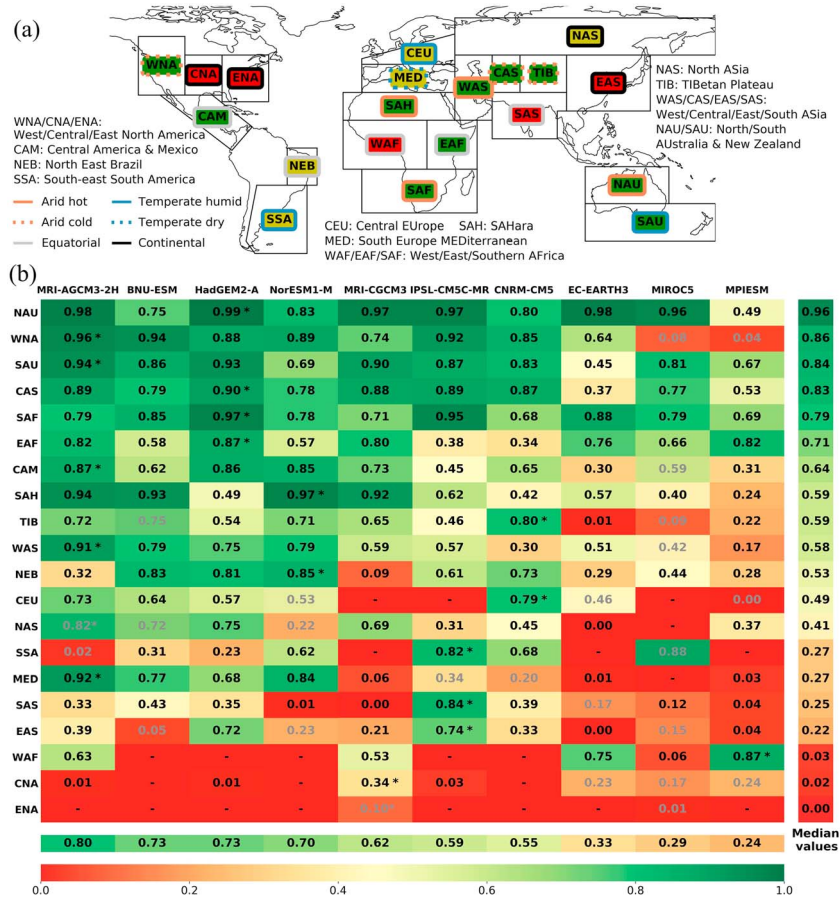
CEU exhibits a broader range of hydrological conditions than SAF, with an observed bell-shaped RWR curve (maximum in the seventh decile), consistent with the analysis over a larger European domain by Folwell et al. (2016). The models with the highest values of  $R^2_{\text{obs}}$  (CNRM-CM5 and MRI-AGCM3-2H) also exhibit a strong inverted bell-shaped EF gradient curve; in other words, EF decreases most rapidly in dry spells with intermediate soil wetness, consistent with theory. The models ranked lowest using  $R^2_{\text{obs}}$  (IPSL-CM5C-MR, MIROC5, MPIESM, and MRI-CGCM3) have almost no sensitivity of RWR to rainfall, at odds with observations. They do not simulate relatively strong dry spell decreases in EF at intermediate rainfall, suggesting poor underlying SM-ET relationships in these models for this region. Note that the CNRM-CM5 model performs best here in spite of sampling a disproportionately high fraction of dry spells following dry months (shown by gray bars in Figure 2). This illustrates that our approach isolates the effect of the underlying land model from biases in rainfall forcing by the host ESM.

The  $R^2_{\text{obs}}$  metric of land model performance in simulating observed LST-SM functional relationships is provided for all SREX regions and models in Figure 3. Broadly speaking, the best agreement between observations and models is found in arid climates, while models struggle to simulate RWR behavior in continental climate regions. For the most part, the ranking of the regions in Figure 3 is not directly attributable to a weakening in the correlation between simulated RWR and EF gradient (Figure S3). In the vast majority of regions and models, there is a strong correlation between our proxy (RWR) and the object of our interest (dry spell ET dynamics). The location (in antecedent rainfall space) of the maximum RWR is the main source of discrepancy between models and observations. We note that the lowest ranked models simulate the location of the RWR curve maximum at higher rainfall than in observations (Figure S4). This implies a tendency toward stage III drying in these model regions, which is inconsistent with observed RWR. This finding is in agreement with previous studies reporting that models tend to overestimate the response of ET to soil drying and simulate overly SM-limited conditions in several regions (Berg & Sheffield, 2018; Dirmeyer et al., 2018; Levine et al., 2016; Sippel et al., 2017).

## 5. Summary and Discussion

In this work, we have developed a methodology for evaluating the sensitivity of ET to SM in ESMs across the world. This is based on the dynamics of LST and  $T_a$  during dry spells of 10 days, when SM limitation may affect the daily evolution of ET. The models show that RWR is well correlated with changes in the partition of surface energy between sensible and latent heat during a dry spell (Figure S3). Comparisons between observed and simulated RWR (Figure 3) thus highlight where models are unable to capture a realistic SM-ET relationship. The results show that in general, models capture dry spell ET dynamics well in arid climate zones. The largest differences, both across models and with observations, occur in continental climate zones. Here there is a tendency for dry spell decreases in EF to be too large (or in other words, too strong an influence of stage III drying).

Our approach exploits the direct response of LST to changing hydrological conditions. This has an advantage over previous studies examining the atmospheric response to SM (for example through  $T_a$ ), which brings in



**Figure 3.** (a) Overview of model performance based on the coefficient of determination ( $R^2_{obs}$ ) between observed and modeled relative warming rate per SREX region. Green/red indicates general agreement/disagreement between observations and models. Region box frames are colored by dominant (most frequent) Köppen-Geiger based climatic zones. (b) Heat map displaying  $R^2_{obs}$  for all model regions. Best performing regions (models) are ordered by median  $R^2_{obs}$  vertically (horizontally). The best performing model per region is indicated with “\*.” Gray numbers indicate that the corresponding  $R^2_{EF}$  in Figure S3 is below 0.7. Cells shaded dark red and annotated with “-” indicate that  $R_{obs}$  is negative.

additional factors external to land processes. On the other hand, using our observational measure (RWR) brings its own challenges. While the quality of LST data derived from satellites is high relative to more derived products (such as ET), we also need accurate values of  $T_a$  from reanalysis. We recognize that the method relies on 1- to 2-K differences between 10-day tendencies in surface and air temperatures, which might be difficult to estimate, though our previous work (GE16) showed relatively little sensitivity in RWR to alternative choices of reanalysis data. Another important issue for the regional scale analyses in Figures 2 and 3 is the spatial and temporal sampling of the dry spells. Climatological patterns of both cloudiness and dry spells dictate where and when the data are sampled. Moreover, it is challenging to interpret results from some of the SREX regions when they incorporate such diverse landscapes as tropical forest and semidesert (e.g., WAF).

We interpret absolute differences between modeled and observed RWR with caution. This is partly because of the aforementioned uncertainties in estimating observed RWR. Additionally, modeled RWR is sensitive to properties such as roughness length and bare soil fraction, and their prescription from remotely sensed land cover maps. Biases in these properties will affect the absolute value of RWR but may only weakly affect dry spell ET dynamics. For these reasons, we consider how RWR varies with antecedent rainfall to be a more robust measure. Finally, our methodology is currently restricted to 10-day dry spells out of a pragmatic choice of sampling sufficient events globally. Land schemes may perform well at this time scale but poorly at longer time scales relevant for seasonal drought (Ukkola et al., 2016).

It is beyond the scope of this paper to provide more than an overview of the approach and summarize model results. We can speculate that simulated dry spell ET dynamics may be poor in CNA, ENA, EAS, and SAS because of their extensive cropland. Many land models lack a realistic representation of the impact of agriculture (particularly when irrigated) on the surface energy balance. The inclusion of groundwater would also suppress simulated ET sensitivity to antecedent rainfall. We find that no single model performs well in all regions. On the other hand, the regional median  $R^2_{\text{obs}}$  used to rank models in Figure 3 clearly distinguishes between good and bad performance against our metric globally, though without further analysis, we would caution against overinterpretation.

Notwithstanding some of the issues highlighted above, we think that this approach provides a valuable new tool for modelers seeking to evaluate and upgrade the description of land surface processes (Balsamo et al., 2018). Considering the growing body of evidence that in many regions ESMs tend to overestimate SM-limited ET conditions (Dirmeyer et al., 2018; Levine et al., 2016; Sippel et al., 2017), our method can make a relevant contribution to identify actual water-limited conditions and surface heat release and help to constrain heat extremes projections. We provide the code and aggregated observations for developers to benchmark their individual models. With the advent of CMIP6, which will see subdaily LST data archived from many more models, we hope to provide a more complete assessment of these state-of-the-art models and continue to refine our methodology to identify model shortcomings.

#### Acknowledgments

This research was funded under the U.K. NERC e-stress project (NE/K015990/1). Additional support was provided by IMPALA (NE/M017230/1) and European Commission grant agreement 282673 (EMBRACE), and through U.K. NERC support of the National Centre for Earth Observation. Observational data used in the study are available from <http://reverb.echo.nasa.gov/> (MODIS), <http://due.esrin.esa.int/> (ERA-Interim), <http://disc.sci.gsfc.nasa.gov> (TRMM), <http://www.cpc.ncep.noaa.gov/> (CMORPH), and <http://chrs.web.uci.edu/> (PERSIANN). Model data are available from <http://pcmdi9.llnl.gov/> (CMIP5) and <http://www.embrace-project.eu> (EMBRACE). We gratefully acknowledge the providers of climate model data in the CMIP5 archive and EMBRACE partners, in particular, Gill Martin from Met Office (United Kingdom), Frederique Cheruy from LMD (France), and Klaus Wyser from SMHI (Sweden), for providing outputs from additional simulations. The code and aggregated observations are available from GitHub website ([https://github.com/ppharris/dry\\_spell\\_rwr](https://github.com/ppharris/dry_spell_rwr)).

#### References

- Balsamo, G., Agustí-Parareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., et al. (2018). Satellite and in situ observations for advancing global Earth surface modelling: A Review. *Remote Sensing*, *10*(12). <https://doi.org/10.3390/rs10122038>
- Berg, A., & Sheffield, J. (2018). Soil moisture-evapotranspiration coupling in CMIP5 models: Relationship with simulated climate and projections. *Journal of Climate*, *31*(12), 4865–4878. <https://doi.org/10.1175/jcli-d-17-0757.1>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. <https://doi.org/10.1002/qj.828>
- Dirmeyer, P. A. (2011). The terrestrial segment of soil moisture-climate coupling. *Geophysical Research Letters*, *38*, L16702. <https://doi.org/10.1029/2011GL048268>
- Dirmeyer, P. A., Chen, L., Wu, J., Shin, C.-S., Huang, B., Cash, B. A., et al. (2018). Verification of land-atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations. *Journal of Hydrometeorology*, *19*(2), 375–392. <https://doi.org/10.1175/jhm-d-17-0152.1>
- Folwell, S. S., Harris, P. P., & Taylor, C. M. (2016). Large-scale surface responses during European dry spells diagnosed from land surface temperature. *Journal of Hydrometeorology*, *17*(3), 975–993. <https://doi.org/10.1175/jhm-d-15-0064.1>
- Gallego-Elvira, B., Taylor, C. M., Harris, P. P., Ghent, D., Veal, K. L., & Folwell, S. S. (2016). Global observational diagnosis of soil moisture control on the land surface energy balance. *Geophysical Research Letters*, *43*, 2623–2631. <https://doi.org/10.1002/2016GL068178>
- Gevaert, A. I., Miralles, D. G., de Jeu, R. A. M., Schellekens, J., & Dolman, A. J. (2018). Soil moisture-temperature coupling in a set of land surface models. *Journal of Geophysical Research: Atmospheres*, *123*, 1481–1498. <https://doi.org/10.1002/2017JD027346>
- Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., et al. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. <https://doi.org/10.1175/jhm560.1>
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive-microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2)
- Koster, R. D., Dirmeyer, P. A., Guo, Z. C., Bonan, G., Chan, E., Cox, P., et al., & GLACE Team (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. <https://doi.org/10.1126/science.1100217>
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. <https://doi.org/10.1127/0941-2948/2006/0130>
- Levine, P. A., Randerson, J. T., Swenson, S. C., & Lawrence, D. M. (2016). Evaluating the strength of the land-atmosphere moisture feedback in Earth system models using satellite observations. *Hydrology and Earth System Sciences*, *20*(12), 4837–4856. <https://doi.org/10.5194/hess-20-4837-2016>
- Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., & de Arellano, J. V.-G. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geoscience*, *7*(5), 345–349. <https://doi.org/10.1038/ngeo2141>
- Miralles, D. G., van den Berg, M. J., Teuling, A. J., & de Jeu, R. A. M. (2012). Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters*, *39*, L21707. <https://doi.org/10.1029/2012GL053703>
- Mueller, B., & Seneviratne, S. I. (2012). Hot days induced by precipitation deficits at the global scale. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12,398–12,403. <https://doi.org/10.1073/pnas.1204330109>
- Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, *41*, 128–134. <https://doi.org/10.1002/2013GL058055>
- Panwar, A., Kleidon, A., & Renner, M. (2019). Do surface and air temperatures contain similar imprints of evaporative conditions? *Geophysical Research Letters*, *46*, 3802–3809. <https://doi.org/10.1029/2019GL082248>
- Schwingshackl, C., Hirschi, M., & Seneviratne, S. I. (2017). Quantifying spatiotemporal variations of soil moisture control on surface energy balance and near-surface air temperature. *Journal of Climate*, *30*(18), 7105–7124. <https://doi.org/10.1175/jcli-d-16-0727.1>



- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4), 125–161. <https://doi.org/10.1016/j.earscirev.2010.02.004>
- Seneviratne, S. I., Luethi, D., Litschi, M., & Schaer, C. (2006). Land-atmosphere coupling and climate change in Europe. *Nature*, 443(7108), 205–209. <https://doi.org/10.1038/nature05095>
- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In *Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)* (pp. 109–230). Cambridge, UK and New York: Cambridge University Press.
- Seneviratne, S. I., Wilhelm, M., Stanelle, T., van den Hurk, B., Hagemann, S., Berg, A., et al. (2013). Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophysical Research Letters*, 40, 5212–5217. <https://doi.org/10.1002/grl.50956>
- Sippel, S., Zscheischler, J., Mahecha, M. D., Orth, R., Reichstein, M., Vogel, M., & Seneviratne, S. I. (2017). Refining multi-model projections of temperature extremes by evaluation against land-atmosphere coupling diagnostics. *Earth System Dynamics*, 8(2), 387–403. <https://doi.org/10.5194/esd-8-387-2017>
- Sorooshian, S., Hsu, K. L., Imam, B., & Hong, Y. (2008). Global precipitation estimation from satellite image using artificial neural networks. In *Hydrological Modelling in Arid and Semi-Arid Areas* (pp. 21–27). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511535734.003>
- Taylor, C. M., de Jeu, R. A. M., Guichard, F., Harris, P. P., & Dorigo, W. A. (2012). Afternoon rain more likely over drier soils. *Nature*, 489(7416), 423–426. <https://doi.org/10.1038/nature11377>
- Teuling, A. J. (2018). Climate hydrology: A hot future for European droughts. *Nature Climate Change*, 8(5), 364–365. <https://doi.org/10.1038/s41558-018-0154-5>
- Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., et al. (2009). A regional perspective on trends in continental evaporation. *Geophysical Research Letters*, 36, L02404. <https://doi.org/10.1029/2008GL036584>
- Teuling, A. J., Seneviratne, S. I., Stockli, R., Reichstein, M., Moors, E., Ciais, P., et al. (2010). Contrasting response of European forest and grassland energy exchange to heatwaves. *Nature Geoscience*, 3(10), 722–727. <https://doi.org/10.1038/ngeo950>
- Teuling, A. J., Seneviratne, S. I., Williams, C., & Troch, P. A. (2006). Observed timescales of evapotranspiration response to soil moisture. *Geophysical Research Letters*, 33, L23403. <https://doi.org/10.1029/2006GL028178>
- Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., et al. (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environmental Research Letters*, 11(10). <https://doi.org/10.1088/1748-9326/11/10/104012>
- Ukkola, A. M., Pitman, A. J., Donat, M. G., De Kauwe, M. G., & Angelil, O. (2018). Evaluating the contribution of land-atmosphere coupling to heat extremes in CMIP5 models. *Geophysical Research Letters*, 45, 9003–9012. <https://doi.org/10.1029/2018GL079102>
- Zscheischler, J., Orth, R., & Seneviratne, S. I. (2015). A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophysical Research Letters*, 42, 9816–9824. <https://doi.org/10.1002/2015GL066563>