Earth System
Dynamics

Open Access

Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics

Sebastian Sippel[1,2]**, Jakob Zscheischler**[2]**, Miguel D. Mahecha**[1]**, Rene Orth**[2]**, Markus Reichstein**[1]**,
Martha Vogel**[2]**, and Sonia I. Seneviratne**[2]

[1]Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany
[2]Institute for Atmospheric and Climate Science, ETH Zürich, 8075 Zurich, Switzerland

Correspondence to: Sebastian Sippel (ssippel@bgc-jena.mpg.de)

Abstract. The Earth's land surface and the atmosphere are strongly interlinked through the exchange of energy and matter. This coupled behaviour causes various land–atmosphere feedbacks, and an insufficient understanding of these feedbacks contributes to uncertain global climate model projections. For example, a crucial role of the land surface in exacerbating summer heat waves in midlatitude regions has been identified empirically for high-impact heat waves, but individual climate models differ widely in their respective representation of land–atmosphere coupling. Here, we compile an ensemble of 54 combinations of observations-based temperature (T) and evapotranspiration (ET) benchmarking datasets and investigate coincidences of T anomalies with ET anomalies as a proxy for land–atmosphere interactions during periods of anomalously warm temperatures. First, we demonstrate that a large fraction of state-of-the-art climate models from the Coupled Model Intercomparison Project (CMIP5) archive produces systematically too frequent coincidences of high T anomalies with negative ET anomalies in midlatitude regions during the warm season and in several tropical regions year-round. These coincidences (high T, low ET) are closely related to the representation of temperature variability and extremes across the multi-model ensemble. Second, we derive a land-coupling constraint based on the spread of the T–ET datasets and consequently retain only a subset of CMIP5 models that produce a land-coupling behaviour that is compatible with these benchmark estimates. The constrained multi-model simulations exhibit more realistic temperature extremes of reduced magnitude in present climate in regions where models show substantial spread in T–ET coupling, i.e. biases in the model ensemble are consistently reduced. Also the multi-model simulations for the coming decades display decreased absolute temperature extremes in the constrained ensemble. On the other hand, the differences between projected and present-day climate extremes are affected to a lesser extent by the applied constraint, i.e. projected changes are reduced locally by around 0.5 to 1 °C – but this remains a local effect in regions that are highly sensitive to land–atmosphere coupling. In summary, our approach offers a physically consistent, diagnostic-based avenue to evaluate multi-model ensembles and subsequently reduce model biases in simulated and projected extreme temperatures.

1   Introduction

The exchange of matter and energy between the land surface and the atmosphere is a crucial feature of the Earth's climate (Seneviratne et al., 2010b; Bonan, 2015; van den Hurk et al., 2016). On one hand, the atmosphere exerts a key influence on land surface processes such as vegetation growth by supplying light, water, and carbon dioxide (Köppen, 1900). On the other hand, the land surface feeds back to the atmosphere, for example through the partitioning of energy into latent and sensible heat fluxes or by modifying land surface properties, thus implying a direct link to near-surface climate (Koster et al., 2004; Seneviratne et al., 2010b). Conceptually, coupling between the atmosphere and the land surface is often classified into two qualitatively different regimes, a "energy-limited" and a "water-limited" regime (Seneviratne et al., 2010b): in the wet (energy-limited) regime, the land surface is largely controlled by the atmosphere through radiation (see conceptual Fig. 1a and b), implying a positive association between near-surface temperature (T) and evapotranspiration (ET). In contrast, in a dry, water-limited state, land controls near-surface climate through a lack of soil moisture, and a corresponding reduction in evapotranspiration and latent cooling (see conceptual Fig. 1a and b) with a negative association between T and ET. Therefore, the state of the land surface and land–atmosphere feedbacks modulate and amplify climatic extreme events such as heat waves in midlatitude regions (Seneviratne et al., 2006; Fischer et al., 2007; Hirschi et al., 2011; Whan et al., 2015; Hauser et al., 2016). An understanding of these feedbacks might yield improved seasonal predictability of extremes (Quesada et al., 2012) and could help to constrain and better predict model-simulated present and future climate variability in these regions (Seneviratne et al., 2006, 2013; Lorenz et al., 2012; Dirmeyer et al., 2013; van den Hurk et al., 2016; Davin et al., 2016).

However, at present large uncertainties and methodological inconsistencies prevail in both understanding and quantification of land–atmosphere coupling on various spatial and temporal scales, which relate to

i.   scarcity of accurate observational products of soil moisture or evapotranspiration on large spatiotemporal scales and relatively short observational periods (Seneviratne et al., 2010b),

ii. the metrics and variables used to quantify land–atmosphere coupling differ widely in the variables they address (Seneviratne et al., 2010b) and in emphasizing either the whole distribution (Dirmeyer, 2011; Lorenz et al., 2012; Miralles et al., 2012) or the tails of relevant variables (Zscheischler et al., 2015).

As a consequence, uncertainties and methodological inconsistencies contribute to a greatly diverging representation of land–atmosphere coupling in state-of-the-art climate models

(Koster et al., 2004; Boé and Terray, 2008; see also Fig. 1a and b for a simple conceptual example), and further contribute to uncertainties related to projected increases in summer temperature variability in the 21st century in midlatitude regions (Seneviratne et al., 2006; Dirmeyer et al., 2013). In this context, it has been noted that accurate simulations of temperature variability and extremes require a realistic representation of land–atmosphere interactions (Seneviratne et al., 2006; Fischer et al., 2012; Bellprat et al., 2013). In other words, biases in temperature variability and extremes might in part stem from an unrealistic representation of land–atmosphere interactions (Fischer et al., 2012; Lorenz et al., 2012; Davin et al., 2016), likely leading to temperature-dependent biases in multi-model ensembles (Boberg and Christensen, 2012; Bellprat et al., 2013).

A model evaluation focus on interpretable land–atmosphere coupling diagnostics might serve as a complementary strategy to traditional model validation and testing (Seneviratne et al., 2010a; Santanello et al., 2010; Mueller et al., 2011b; Mueller and Seneviratne, 2014). Hence, this approach is intended towards testing and understanding the spread and physical consistency in simulated relationships in state-of-the-art multi-model ensembles (e.g. the Coupled Model Intercomparison Project, CMIP5; Taylor et al., 2012) against available observations-based datasets. For example, in the context of land–atmosphere coupling, earlier studies used bivariate correlation- or regression-based metrics to test and evaluate coupling behaviour (Hirschi et al., 2011; Lorenz et al., 2012). Conceptually, the notion of "diagnostic-based model evaluation" as discussed here is consistent with "pattern-oriented model evaluation" (Grimm and Railsback, 2012; Reichstein et al., 2011) – the latter being applied in the context of evaluating simulated and observed patterns on multiple scales in a data-driven way (e.g. in the context of ecosystem carbon turnover times; Carvalhais et al., 2014).

In the context of extracting credible and relevant information from large (multi-)model ensembles, weighting or selecting models based on observations-based constraints has become increasingly popular recently (Tebaldi and Knutti, 2007; Knutti, 2010), as a priori model ensembles might be seen as a somewhat arbitrary collection of model runs (or "ensembles of opportunity"). For example, empirical and/or physics-based criteria have been used to constrain snow–albedo feedbacks (Hall and Qu, 2006), constrain carbon cycle projections (Cox et al., 2013; Wenzel et al., 2014; Mystakidis et al., 2016), or in the context of refining precipitation projections (Orth et al., 2016). Moreover, empirical diagnostics are applied to select models for event attribution analyses (Perkins et al., 2007; King et al., 2016; Otto et al., 2015) and analyses of drought projections based on model performance (Van Huijgevoort et al., 2014) or to resample large initial-condition ensembles to alleviate biases without distorting the multivariate structure of climate model output (Sippel et al., 2016b). In the context of land–atmosphere coupling, Fischer et al. (2012) and Stegehuis et al. (2013) have constrained a re-
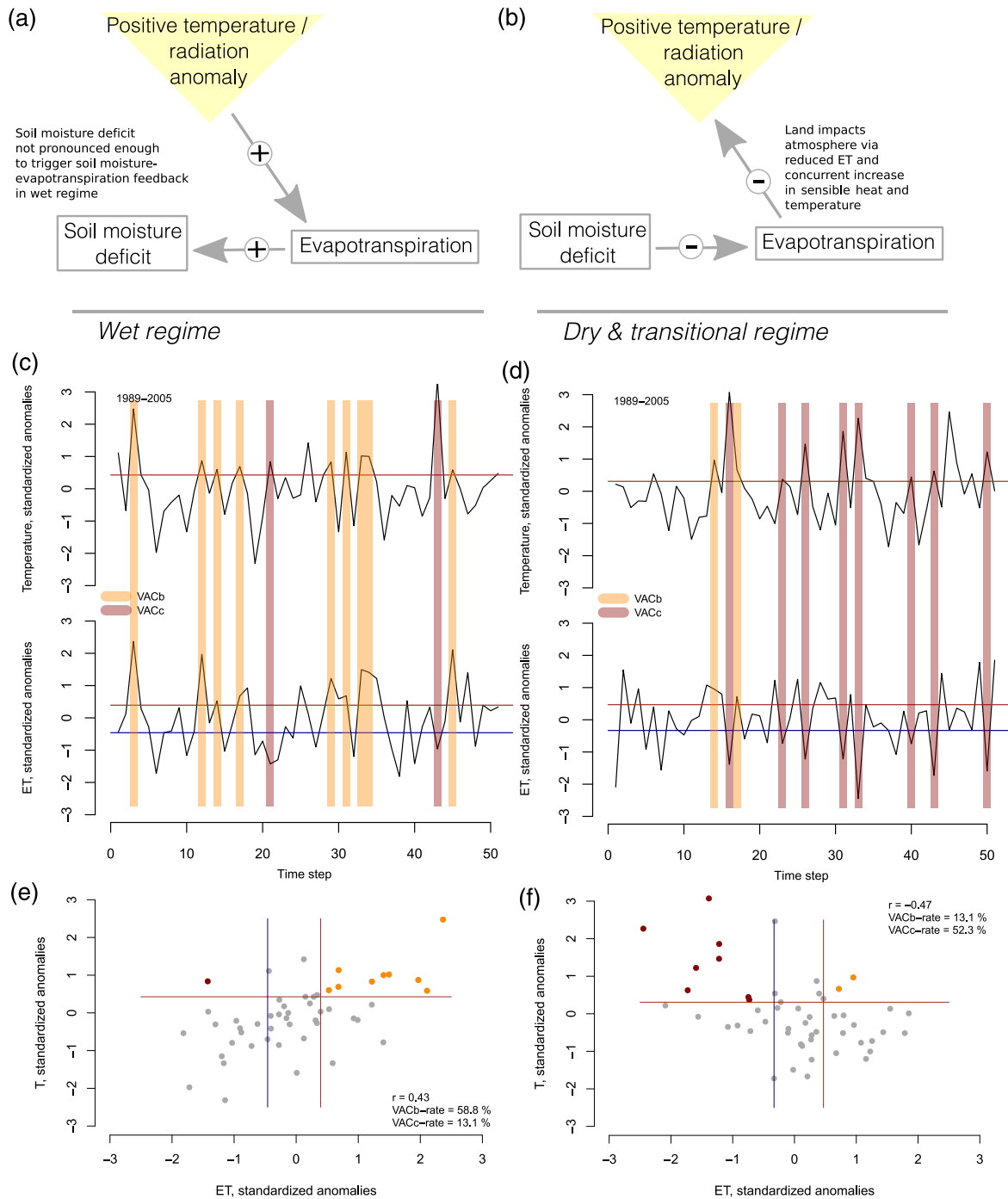
Figure 1. Illustration of qualitatively contrasting warm season temperature–evapotranspiration (T–ET) coupling in global climate models. **(a, b)** Conceptual illustration of T–ET coupling in **(a)** wet and **(b)** dry and transitional regimes. In wet regimes T and ET are positively associated (atmosphere impacts land), while in dry and transitional regimes T and ET are negatively associated due to soil moisture feedbacks (i.e. land impacts atmosphere via reduced ET and concurrent increases in sensible heat and T). **(c)–(f)** Different CMIP5 models show contrasting T–ET coupling behaviour in a midlatitude region in summer (central Europe, spatial average, JJA, 1989–2005): **(c, e)** NorESM1-M produces predominantly wet regimes, i.e. a positive T–ET coupling, while **(d, f)** ACCESS1-3 produces predominantly dry regimes (negative T–ET coupling), illustrated as time series **(c, d)** and in the T–ET plane **(e, f)**. Dark red lines in **(c)–(f)** indicate upper threshold (th_{upper}^{T}) for T and ET; dark blue lines indicate lower threshold (th_{lower}^{T}) (70th and 30th percentile in each individual time series).

gional model ensemble over Europe using present-day inter-annual variability in summer temperature, and observations-based estimates of summer sensible heat fluxes. However, these studies came to somewhat conflicting results with respect to the obtained change in warming projections, which was probably due to the underlying choices of datasets to obtain the constraints (Stegehuis et al., 2013). Hence, care is needed in that these practices might not necessarily translate into improved future climate projections or reduced uncertainties. That is because the selection of relevant metrics is clearly not trivial but subjective and because good model performance with regard to any given metric does not translate directly into (more) reliable projections (Knutti, 2008).

Therefore, the starting point for the present analysis – in the sense of being necessary, but not sufficient to ensure the reliability of future climate projections – is that physically motivated, observations-based diagnostics might offer

1. a link to identify and interpret relevant processes across multiple models (i.e. model evaluation) and

2. to reduce biases by focusing the interpretation of multi-model ensembles on models that are "right for the right reasons". Most notably climate impacts, including extremes, typically depend on the multivariate structure of climate variables, where simple univariate statistical bias correction methods are prone to failure (Ehret et al., 2012; Cannon, 2016).

In this study, we first evaluate land–atmosphere coupling in state-of-the-art global climate models from the CMIP5 archive and a large ensemble of observations-based ET datasets (Mueller et al., 2013) that has been compiled to address the aforementioned uncertainties in land–atmosphere coupling. In our analyses a land–atmosphere coupling metric that is based on coincidences of temperature and evapotranspiration anomalies is applied. The idea behind a coincidence metric as opposed to a traditional univariate evaluation of model-simulated ET fluxes or temperature is that it is insensitive to biases in the simulated means or variances and thus focusses only on an abstract property of the data, namely the bivariate dependence structure of T and ET. Secondly, we derive a model constraint based on the physically motivated land-coupling diagnostic and the ensemble of benchmarking datasets in order to explore the implications of a reduced ensemble but with land–atmosphere coupling that is *within the range* of the benchmarking datasets.

2 Data and methods

2.1 Datasets for T–ET coupling analysis and model evaluation

2.1.1 Global temperature and evapotranspiration datasets

In order to evaluate T–ET coupling in global climate models, an ensemble of 18 gridded ET estimates, taken from the LandFlux-EVAL multi-dataset synthesis project (Mueller et al., 2013), are combined with three different observations-based and reanalysis-driven temperature datasets, yielding in total 54 T–ET combinations (see Table 1). T–ET coincidence rates are calculated from each of those 54 combinations to evaluate and constrain the multi-model ensemble of global climate models (Sect. 3). The ensemble of ET reference datasets has been generated by combining a wide range of different ET estimates, consisting of five diagnostic (based on remote-sensing or in situ observations) products, five land surface models driven by observed climate forcing, and four reanalysis products (Mueller et al., 2013). The three temperature datasets are based on one observational product (Climate Research Unit dataset, Harris et al., 2014) and two reanalysis products (ERA-Interim reanalysis (ERAI; Dee et al., 2011), and Climate Forecast System Reanalysis (CFSR; Saha et al., 2010); see Table 1 for details). The large number of T–ET dataset combinations is used in order to take uncertainties in both T and ET datasets into account. We have tested that the spread between individual ET datasets is substantially larger than the spread between individual T datasets (not shown). This indicates that the largest source of uncertainty stems from the choice of ET dataset, and therefore we consider only three different T datasets. Each of the 54 T–ET dataset combinations (denoted as "T–ET coupling benchmarks" in the remainder of the paper) is consistently derived from observations and thus can be expected to represent relevant features in T–ET coupling under different assumptions that underlie diagnostic datasets, reanalyses, and land surface models. Therefore, these datasets represent a very large spread of plausible T–ET coupling estimates, and the spread can be considered as a conservative benchmark for model evaluation (including observational noise, i.e. allowing a wide range of T–ET coupling in models). However, it should be emphasized that the datasets are not independent realizations. Thus, we only use the spread of the T–ET coupling benchmarks, but we do not interpret the probability distribution of dataset combinations.

For the analysis of historical and future simulations of the monthly maximum value of daily maximum temperatures (TXx) in Sect. 3.2 we use ERA-Interim (Dee et al., 2011) as a reference dataset.

Table 1. Datasets used for model evaluation. LSM: land surface model.

Name of dataset	Variable	Type/group	Provider and reference
LandFlux-EVAL[a]	ET	Ensemble median	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of reanalyses	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of LSMs	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of diagnostic datasets	Mueller et al. (2013)
PRUNI[a,b]	ET	Diagnostic	Sheffield et al. (2006, 2010)
MPIBGC[a,b]	ET	Diagnostic	Jung et al. (2011)
CSIRO[a,b]	ET	Diagnostic	Zhang et al. (2010)
GLEAM[a,b], V. 1A	ET	Diagnostic	Miralles et al. (2011a, b)
AWB[a,b]	ET	Diagnostic	Mueller et al. (2011a)
EI-ORCHIDEE[a,b]	ET	LSM	Krinner et al. (2005)
CRU-ORCHIDEE[a,b]	ET	LSM	Krinner et al. (2005)
VIC[a,b]	ET	LSM	Sheffield et al. (2006), Sheffield and Wood (2007)
GL-NOAH-PF[a,b]	ET	LSM	Rodell et al. (2004), Rui and Beaudoing (2016)
MERRA-LAND[a,b]	ET	LSM	Reichle et al. (2011)
ERA-Interim[a,b]	ET	Reanalysis	Dee et al. (2011)
CFSR[a,b]	ET	Reanalysis	Saha et al. (2010)
JRA-25[a,b]	ET	Reanalysis	Onogi et al. (2007)
MERRA[a,b]	ET	Reanalysis	Bosilovich (2008)
CRU-TS3.2[a]	T	Observations	Harris et al. (2014)
ERA-Interim reanalysis[a]	T	Reanalysis	Dee et al. (2011)
CFSR reanalysis[a]	T	Reanalysis	Saha et al. (2010)

[a] All T–ET combinations of marked datasets have been used to derive the ET–T constraint. [b] Original individual datasets that contributed to the LandFlux-EVAL synthesis project (Mueller et al., 2013).

2.1.2 Multi-model ensemble simulations

The Climate Model Intercomparison Project (CMIP5) has been designed to allow for multi-model comparison and evaluation studies (Taylor et al., 2012). Although a large model spread, biases, and uncertainties remain in the ensemble projections (Knutti and Sedláček, 2013), for example with respect to extremes (Sillmann et al., 2013a), the water (Mueller et al., 2011b; Mueller and Seneviratne, 2014), and land carbon cycle (Anav et al., 2013), the archive of standardized scenario-driven model experiments provides one of the main avenues to study climate variability and change (e.g. Stocker et al., 2013), including present and future climate extremes (Sillmann et al., 2013b; Seneviratne et al., 2016). We use one ensemble member from 37 individual models or model variants (Table S1 in the Supplement) to avoid unequal sample sizes in the multi-model ensembles. Furthermore, this choice is made to assess variability in land–atmosphere coupling *across* models because individual ensemble members from the same model show a comparably small spread in land–atmosphere coupling and present-day and future land–atmosphere coupling are highly correlated (Fig. S1 in the Supplement; metric and definition is provided below). This indicates that the large spread between models is dominated by variability *across* models, and thus land–atmosphere coupling is a model-inherent feature on climatological timescales (Figs. S1 and S2; see further discussion below). On shorter (e.g. annual or seasonal) timescales,

models do indeed show substantial variability in their land–atmosphere coupling (Sippel et al., 2016b), which could be used as a constraint in large single-model ensembles, but this is beyond the scope of the present study.

2.1.3 Data processing and analysis

All datasets were remapped to a common $2.5° × 2.5°$ spatial resolution for analysis and before computing T–ET coincidences. For model evaluation (Sect. 3.1), all computations and analyses are performed on a monthly temporal resolution and are restricted to the time period 1989–2005 due to data availability constraints of the ET reference datasets (Mueller et al., 2013). Thus, the reference period for model evaluation corresponds to the last 17 years of the "historical" scenario in CMIP5 models. T–ET coincidences are computed based on monthly deseasonalized and linearly detrended time series of T and ET, and coincidence rates are calculated separately for each individual season. Only land pixels outside of desert regions following the Köppen–Geiger climate classification are considered (Kottek et al., 2006). The model evaluation is conducted based on all individual pixels and additionally on area averages for IPCC-SREX (Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation) regions (IPCC, 2012).

2.2 Diagnostic-based model evaluation using T–ET coupling

2.2.1 The T–ET link and the Vegetation–Atmosphere Coupling (VAC) index

An adequate characterization of the coupling between soil moisture and temperature is key to model evaluation using observations-based datasets. This coupling is often diagnosed by correlation-based metrics such as for example between T and ET, $\rho_{(T,\mathrm{ET})}$ (Seneviratne et al., 2006; Lorenz et al., 2012), or the difference in the covariability in temperature and sensible heat, where the latter is calculated with and without accounting for soil moisture deficits (Miralles et al., 2012). Here, we aim to exploit the T–ET coupling by using a natural extension of $\rho_{(T,\mathrm{ET})}$ that focusses on the tails of T–ET dependencies. Deseasonalized and detrended time series of ET (x_i^{ET}) and T (x_i^T; i denotes the time step) are partitioned into five distinct classes of Vegetation–Atmosphere Coupling (VAC) following Zscheischler et al. (2015), resulting in a time series of discrete events x_i^{VAC}:

$$x_i^{\mathrm{VAC}} = \begin{cases} a, & \text{if} & x_i^T < \mathrm{th}_{\mathrm{lower}}^T & \text{and} & x_i^{\mathrm{ET}} < \mathrm{th}_{\mathrm{lower}}^{\mathrm{ET}}, \\ b, & \text{if} & x_i^T > \mathrm{th}_{\mathrm{upper}}^T & \text{and} & x_i^{\mathrm{ET}} > \mathrm{th}_{\mathrm{upper}}^{\mathrm{ET}}, \\ c, & \text{if} & x_i^T > \mathrm{th}_{\mathrm{upper}}^T & \text{and} & x_i^{\mathrm{ET}} < \mathrm{th}_{\mathrm{lower}}^{\mathrm{ET}}, \\ d, & \text{if} & x_i^T < \mathrm{th}_{\mathrm{lower}}^T & \text{and} & x_i^{\mathrm{ET}} > \mathrm{th}_{\mathrm{upper}}^{\mathrm{ET}}, \\ 0 & \text{otherwise} \end{cases}$$

Event thresholds $\mathrm{th}_{\mathrm{lower}}$ and $\mathrm{th}_{\mathrm{upper}}$ might be chosen relative to the variability in each time series by fixing the probability p to exceed or fall below a threshold through the choice of an appropriate quantile:

$$\Pr\left[X > \mathrm{th}_{\mathrm{upper}}\right] = \Pr\left[X < \mathrm{th}_{\mathrm{lower}}\right] = p. \qquad (1)$$

Taking time series length restrictions into account, we choose the 30th and 70th percentile as lower and upper thresholds in all time series (i.e. such that $\Pr[X < \mathrm{th}_{\mathrm{lower}}] = \Pr[X > \mathrm{th}_{\mathrm{upper}}] = 0.3$). Here, we focus on coincidences of *warm temperature anomalies* ("T events": $x_i^T > \mathrm{th}_{\mathrm{upper}}^T$) with anomalies in ET ("ET events", i.e. either $x_i^{\mathrm{ET}} > \mathrm{th}_{\mathrm{upper}}^{\mathrm{ET}}$ for VAC$_b$ or $x_i^{\mathrm{ET}} < \mathrm{th}_{\mathrm{lower}}^{\mathrm{ET}}$ for VAC$_c$); we derive coincidence rates r_{VAC_b} by counting the number of VAC$_b$ events (see Quiroga et al., 2002, and Donges et al., 2016, for earlier formulations of event coincidence analysis and, e.g., Rammig et al., 2015, and Siegmund et al., 2016, for applications in an ecological context):

$$r_{\mathrm{VAC}_b} = \frac{1}{N_0} \sum_{i=1}^{N} 1_{[b]}\left(x_i^{\mathrm{VAC}}\right). \qquad (2)$$

Here, $1_A(x)$ is the indicator function, defined as $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise; N denotes the length of the time series. Hence, we simply count coincidences of T and ET in a given category (e.g. positive T *and* positive ET for VAC$_b$) to get the average coincidence rate (r_{VAC_b}). N_0 acts as a normalization constant and is chosen in our study such that

$0 \leq r_{\mathrm{VAC}_b} \leq 1$, i.e. we normalize with the total number of T events, $N_0 = \sum_{i=1}^{N} 1_{[x^T > \mathrm{th}_{\mathrm{upper}}^T]}(x_i^T)$. Hence, if all (or none) of the T events in the time series coincided with ET events, then the average coincidence rates would be given by $r_{\mathrm{VAC}_b} = 1$ (or $r_{\mathrm{VAC}_b} = 0$). For independent time series, i.e. no coupling, r_{VAC_b} would approximate the occurrence rate of ET events in the time series (defined for VAC$_b$) that is governed by the chosen threshold, i.e. $r_{\mathrm{VAC}_b} = \frac{1}{N} \sum_{i=1}^{N} 1_{[x_i^{\mathrm{ET}} > \mathrm{th}_{\mathrm{upper}}^{\mathrm{ET}}]}(x_i^{\mathrm{ET}})$ (hence, $r_{\mathrm{VAC}_b} \approx 0.3$ in our case). Coincidence rates r_{VAC_c} follow equivalently by replacing VAC$_b$ with VAC$_c$ in Eq. (2). We compute r_{VAC_b} and r_{VAC_c} for all seasons but with an emphasis on the warmest season of the year. In this study, significance of coincidence rates is established by randomly permuting one time series with respect to the other 100 times. Hence, VAC rates from models or observations-based benchmarks that fall outside the 5th to 95th percentile range of the VAC rates obtained from randomly permuted time series are significantly different from independent data at the 0.1 level.

In other words, r_{VAC_b} gives the fraction of the highest 30 % of temperatures that coincide with the highest 30 % of ET (i.e. occurrence rate of energy-limited regimes), while r_{VAC_c} denotes the fraction of the highest 30 % temperatures that correspond with the lowest 30 % ET (i.e. occurrence rate of water-limited regimes). Figure 1c and d show a simple example of monthly time series of T and ET simulated from two CMIP5 models and occurrences of VAC$_b$ and VAC$_c$ are highlighted, and Fig. 1e and f show the correlation of T and ET. Note that for the same region (area average over central Europe, CEU) and time of the year (monthly data for June, July, and August), one model produces predominantly energy-limited regimes (VAC$_b$; Fig. 1c and e and compare to conceptual illustration in Fig. 1a), whereas the other model produces predominantly water-limited regimes (VAC$_c$; Fig. 1d and f and concept in Fig. 1b).

We abbreviate the average occurrence rates r_{VAC_b} and r_{VAC_c} as VAC$_b$ and VAC$_c$ for convenience in the remainder of the paper. In comparison to more traditional coupling metrics, such as e.g. $\rho_{(T,\mathrm{ET})}$, VAC might be expected to yield similar results on very long timescales, whereas on shorter timescales the VAC index picks up non-linearities in the tails (e.g. during warm temperature anomalies). On the monthly timescale (as used in this study), VAC$_b$ and VAC$_c$ detect distinct non-linearities in models and observations in summer T–ET coupling, e.g. in CEU: Fig. S3 shows that, by correlating VAC$_b$ with VAC$_c$ derived from individual models, observations-based benchmarks, and from a two-dimensional Gaussian distribution, VAC$_b$ and VAC$_c$ rates in models and observations-based benchmarks exceed those that would be expected in random data. This deviation indicates that the warm tail is indeed different from the remainder of the distribution (we observe no such deviation for the cold tail; Fig. S3), and hence an evaluation metric that fo-

cuses on the tail such as the VAC index is indeed useful for our present purpose. In addition to the main text, the model evaluation is presented for $\rho_{(T,\text{ET})}$ to demonstrate robustness to the chosen methodological approach (Fig. S4) and for the VAC index using a 90th percentile threshold (Fig. S4). Both alternatives show qualitatively similar results (see Sect. 3).

2.2.2 A constraint on T–ET coupling in multi-model ensembles

In general, a constraint links an observations-based diagnostic with a key model output variable across multiple models (Cox et al., 2013) and thus can be used to reduce model uncertainties and spread. Here, we derive a T–ET coupling constraint as the uncertainty range from the 54 combinations of T–ET benchmarking datasets. A Gaussian kernel with reliable data-based bandwidth selection (Sheather and Jones, 1991) is fitted over all 54 1989–2005 coincidence rates (r_{VAC_c}) for each meteorological season and pixel (and each SREX region average). Throughout this paper, the 5th to 95th percentile range of the fitted Gaussian kernels is taken as the plausible range of observations, and the reduced (constrained) ensemble of CMIP5 simulations is obtained by retaining only those CMIP5 models that simulate T–ET coincidences that fall within this range of observational uncertainty.

3 Results and discussion

In this section, we first evaluate land coupling in CMIP5 models explicitly against an observations-based ensemble of T–ET combinations and explore the link to temperature variability and extremes (Sect. 3.1). All model evaluation results are presented globally and exemplarily for central Europe (CEU) as a region where global models and observations differ widely. Subsequently, we constrain the ensemble of CMIP5 models using each model's land coupling as diagnosed through the VAC_c index and discuss implications for biases in simulated present-day temperature extremes and warming projections (Sect. 3.2).

3.1 Evaluation of land–atmosphere coupling in CMIP5 models and the link to temperature variability and extremes

3.1.1 Evaluation of T–ET coupling in CMIP5 models

Models and observations-based datasets show a relatively large spread in their representation of T–ET coupling, as expressed exemplarily in CEU through both VAC_b and VAC_c across various seasons (Fig. 2a and b) or diagnosed through more traditional coupling metrics such as $\rho_{(T,\text{ET})}$ (Fig. S4). Individual models indicate pronounced qualitative differences in the warm season, where some models point to energy-limited conditions, whereas others indicate predominantly water-limited ones (Figs. 2a and b and 1, for an illus-

trative example). Observations-based T–ET datasets agree qualitatively, i.e. indicating energy-limited to neutral conditions in the CEU example, thus implying an overestimation of water-limited regimes in CEU in roughly 50 % of CMIP5 models (Fig. 2).

This pattern holds across most regions of the globe, as many CMIP5 models consistently overestimate occurrences of VAC_c regimes (and correspondingly underestimate VAC_b occurrences) in the warm season of the year (Fig. 2c and d; see Fig. S5 for a definition of the warm season in each pixel). In midlatitude and several tropical regions (e.g. central North America, central Europe, the Amazon, India, parts of Africa), more than 25 % and up to 50 % of CMIP5 models lie outside the observational range (Fig. 2d). These discrepancies hold also if metrics that emphasize the whole distribution ($\rho_{(T,\text{ET})}$) or more extreme parts of the tail (VAC based on a 90th percentile threshold) are used for model evaluation (Fig. S4; results for individual seasons are presented for VAC_c and VAC_b in Figs. S6 and S7, respectively). Moreover, the spread between the individual models' representation of land–atmosphere coupling strongly exceeds the spread in observational datasets, although different diagnostic, reanalyses, and land surface model datasets are included in the observations-based ensemble (Fig. 2e for CMIP5 model spread and Fig. 2f for spread in observations-based benchmark datasets).

Furthermore, the models' land–atmosphere coupling, as diagnosed here through the VAC index, is a highly model-inherent feature, as different model variants or ensemble members from the same model generally lie relatively close to each other (Figs. S1 and S2). However, model-specific signatures of model output are not unusual, as diagnosed before for spatial patterns of temperature and precipitation (Knutti et al., 2013) or the statistical information content in carbon fluxes (Sippel et al., 2016a). Furthermore, present-day land–atmosphere coupling is strongly related to future land–atmosphere coupling in the individual models (Fig. S1). A detailed overview of VAC_c coupling in individual models and ensemble members relative to the benchmark datasets for central Europe and central North America is presented in Figs. S1 and S2. Despite regionally pronounced qualitative discrepancies, it should be noted that on a global scale, the distribution of water-limited and energy-limited patterns in models and observations agrees qualitatively (Fig. S8). Likewise, the findings of climatologically too pronounced water-limited regimes in individual models with regard to observations does not exclude the possibility of future changes in the coupling strength in transitional regions (Seneviratne et al., 2006) or of strong water limitations during extreme events in the real world (Miralles et al., 2012; Whan et al., 2015). To this end, an evaluation of the year-to-year variability in the coupling behaviour in larger ensembles of individual models, including very rare events, could constitute a topic for further study, as this study was restricted to relatively moderate events in a 16-year period (70th per-
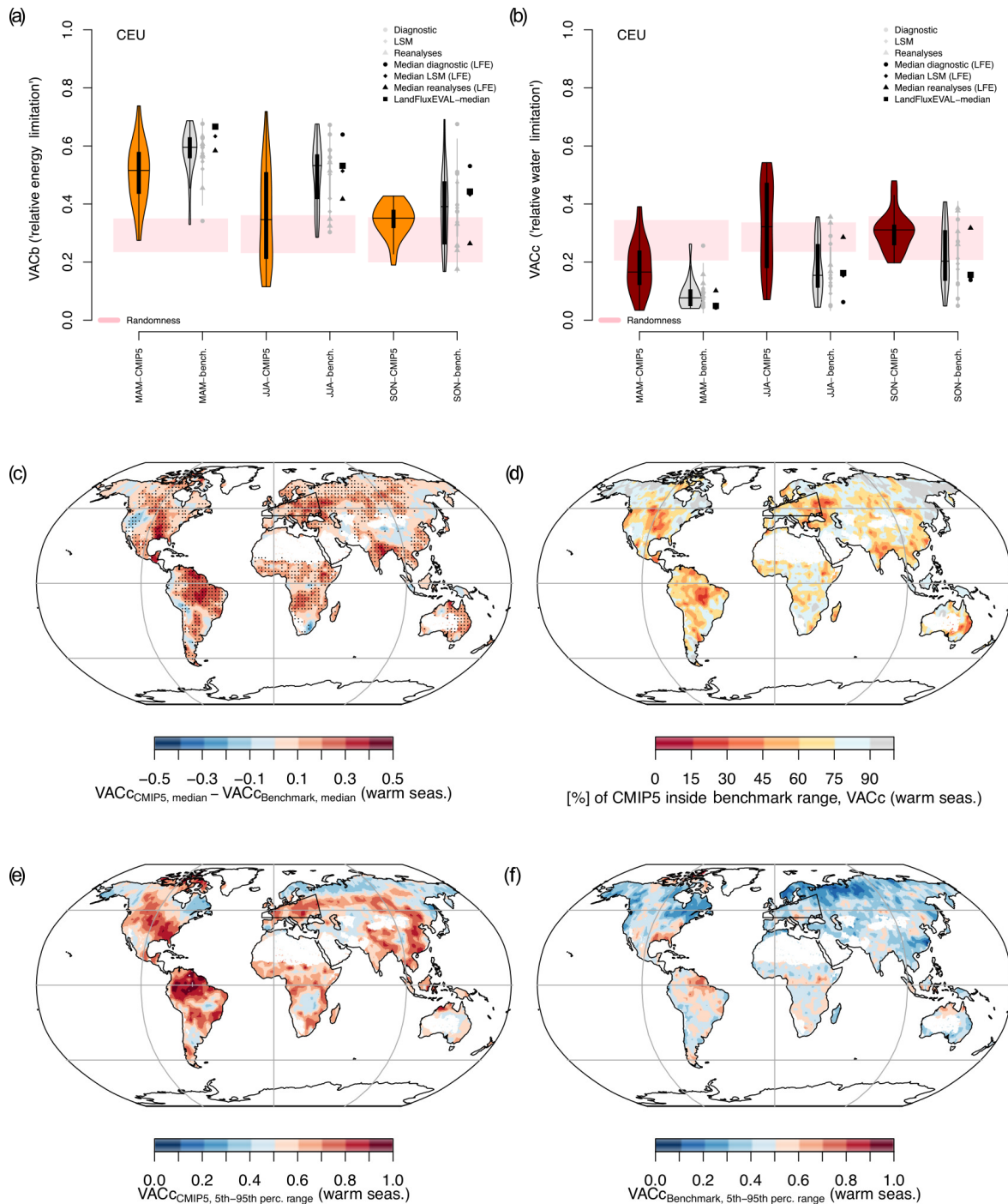
Figure 2. Evaluation of T–ET coupling in global climate models. **(a, b)** VACb and VACc coupling in the CMIP5 climate model ensemble and observations-based benchmarking datasets in central Europe (CEU, 1989–2005, area average) with systematic warm season differences (circles, diamonds, and triangles indicate diagnostic, land surface models, and reanalyses reference datasets, respectively). Randomness indicates the 5th to 95th percentile range obtained by randomly permutating both time series with respect to the other ($N = 100$ times) to obtain independent data. **(c)** Difference in the VACc median of the CMIP5 ensemble and benchmarking datasets. **(d)** Fraction of CMIP5 models that are inside the 5th-95th percentile spread of the benchmarking datasets. **(e, f)** Range of VACc occurrences (5th to 95th percentile range) in CMIP5 models **(e)** and in the ensemble of observations **(f)**.
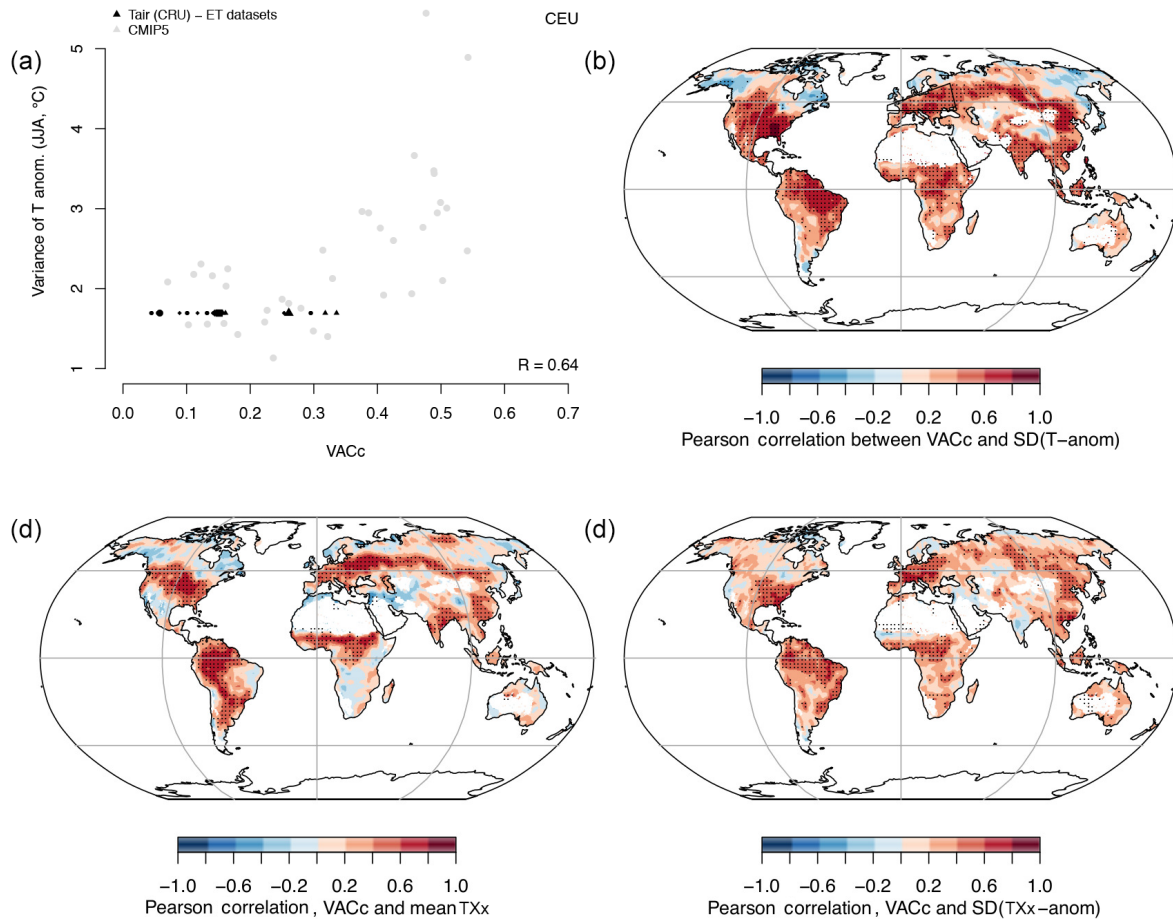
Figure 3. (a, b) Relationship between model-specific T–ET coupling (expressed through VACc) and model-simulated variability in monthly temperature anomalies (JJA) in central Europe **(a)** and globally **(b)**. **(c, d)** Relationship betweeen VACc coupling and mean **(c)** and standard deviation **(d)** of simulated monthly maximum value of daily maximum temperature (TXx) in summer (JJA).

centile threshold for the computation of the VAC index) and one ensemble member per model. Moreover, we also note that observations-based benchmark datasets show systematic (albeit smaller) differences in the representation of land–atmosphere coupling: diagnostic datasets indicate more frequent energy-limited regimes (see, e.g., Fig. 2) and thus differ consistently to generally drier land surface models and reanalysis products, consistent with earlier findings (Santanello et al., 2015).

3.1.2 T–ET coincidences and the link to temperature variability and extremes

The representation of T–ET coupling as diagnosed through the VAC index largely determines the variability in temperatures on monthly and interannual timescales across the CMIP5 multi-model ensemble in CEU (Fig. 3a) and in most regions of the globe except in some subarctic climates (Fig. 3b). Therefore, this relationship is indicative of the strong influence of land–atmosphere coupling on surface climate. This is consistent with previous findings in Europe

in models with and without land–atmosphere interactions (Seneviratne et al., 2006; Fischer and Schär, 2009; Fischer et al., 2012). An important result is that models that produce VAC_c indices within the range of benchmark datasets also produce a realistic near-surface temperature variability, whereas models that fall too frequently into water-limited regimes also overestimate summer temperature variability (Fig. 3a). Moreover, in midlatitude and tropical regions, the state of the land surface is strongly associated with the mean and variability in temperature extremes on the daily timescale in the warmest season (TXx; Fig. 3c and d). The link between the representation of land–atmosphere coupling and simulated temperature extremes and variability in global climate models is consistent with earlier studies, which has been demonstrated for Europe in individual models (Seneviratne et al., 2006; Lorenz et al., 2012; Davin et al., 2016) and in ensembles of regional models (Fischer et al., 2012; Bellprat et al., 2013). Therefore, the relationship between T–ET coincidence rates and temperature extremes might offer an avenue to derive an explicit land–atmosphere coupling con-

straint (the likely root cause of biases) to alleviate biases in temperature variability and extremes in the multi-model CMIP5 ensemble.

3.2 Analysis of constrained multi-model ensemble and implications for future climate projections

3.2.1 A constraint on land–atmosphere coupling in the CMIP5 ensemble

The association between T and ET in the constrained ensemble resembles the observations-based benchmarking datasets in T–ET coupling very well (shown as a bivariate density estimate in Fig. 4a and b for CEU and CNA (central North America), respectively), whereas the unconstrained CMIP5 ensemble produces too many occurrences of VAC_c conditions in both CEU and CNA. Due to the intimate link between land–atmosphere coupling and temperature variability and extremes (see previous section), we expect that the improvement in the representation of land–atmosphere coupling in the constrained ensembles yields a corresponding improvement in the representation of temperature extremes on the daily timescale in coupling-sensitive regions.

Coupling-sensitive regions are prone to warm season biases in climate models (Christensen and Boberg, 2012; Bellprat et al., 2013). In the present analysis, high biases in temperature extremes are indeed prevalent in the original (unconstrained) CMIP5 ensemble in these regions (Fig. 4c and e). For example, the ensemble mean warm season TXx is overestimated by up to 5 °C, and higher biases are detected in the 90th percentile of TXx in CNA, CEU, and the Amazon (all biases in daily variables relative to ERA-Interim; see Fig. 4c and e). In a CMIP5 ensemble constrained by the land–atmosphere coupling metric VAC_c, the representation of temperature extremes is improved in regions prone to coupling-induced biases (Fig. 4d and f); i.e. both mean TXx and the 90th percentile of TXx are significantly reduced. The ensemble mean of present-day temperature extremes in other regions remains unchanged. Moreover, projected future temperature extremes are reduced in the constrained ensemble (Fig. 5), similarly to present-day reductions in regions prone to present-day biases in land–atmosphere coupling. This is illustrated in Fig. 5a for TXx (monthly area averages in summer) in CEU, where the hot end of the original model ensemble is in fact never realized in observed temperatures. The application of the constraint thus not only affects mean TXx but also reduces the spread of the model ensemble (Fig. 5a and b). The reduction in ensemble mean and ensemble spread is retained for the entire 21st century (Fig. 5a and b). Hence, this result reinforces that coupling-related biases are model-inherent features, i.e. models that simulate too many VAC_c occurrences today (and associated high biases in extreme temperatures) are very likely to do so in the future. However, one should keep in mind that the reduction in ensemble mean and spread is confined to coupling-sensitive regions

in CEU, CNA, and to some degree in the Amazon region (Fig. 5c and d).

Our results imply that an accurate representation of land surface processes is crucially relevant for a correct simulation of temperature extremes and more generally for simulated near-surface climate variability. Land–atmosphere coupling is thus an important source of bias in state-of-the-art global climate model simulations. By using an observations-based land–atmosphere coupling diagnostic to constrain the multi-model CMIP5 ensemble, we have shown that biases in extremes in the large ensemble can be alleviated to a certain degree. As bias correction methodologies that take the physical causes of biases into account are still widely lacking (Ehret et al., 2012; Bellprat et al., 2013) and multivariate bias correction methods are currently in development (Cannon, 2016), the identification of models with a *physically plausible* representation of near-surface climate and land–atmosphere interactions on the regional scale might be crucial to extract accurate and relevant information about climate extremes in the context of climatic changes in the 21st century (Mitchell et al., 2016b; Schleussner et al., 2016; Seneviratne et al., 2016). For example, model selection for event attribution studies or a quantification of changes in univariate climate extremes is often based on a statistical performance criterion (Perkins et al., 2007; King et al., 2016; Otto et al., 2015). Our results indicate that these procedures could be further refined through incorporating observations-based diagnostics or constraints in order to analyse model simulations that are indeed right for the right reasons (at least given physics-guided and observations-based relationships). Moreover, the impacts of climate and its extremes, e.g. on human health or ecosystems (Mitchell et al., 2016a; Frank et al., 2015), are often inherently related to multiple climate variables (Ehret et al., 2012; Leonard et al., 2014). Therefore, simple constraints as suggested for instance in the present study might complement more conventional bias correction procedures (e.g. Hempel et al., 2013) to derive physically consistent estimates of climate impacts. This approach appears promising because biases within climate models (i.e. in different variables) and across climate model ensembles are often correlated (e.g. Knutti, 2010; Mueller and Seneviratne, 2014; Sippel et al., 2016b). Hence, beyond soil moisture control on simulated temperature extremes as the present study's focus, related biases in other variables such as warm season precipitation or ET might be similarly relevant in this context. For example, VAC_c occurrences across the CMIP5 ensemble are negatively associated with precipitation and ET in the warm season in midlatitude regions (Fig. S9) – both crucial variables in the water cycle that show pronounced summer low biases in CMIP5 models (Mueller and Seneviratne, 2014). Therefore, a constrained model ensemble with improved land–atmosphere coupling, a likely root cause of biases (Lorenz et al., 2012), might not only improve temperature extremes and variability but additionally might reduce biases in associated variables such as ET or precipitation.
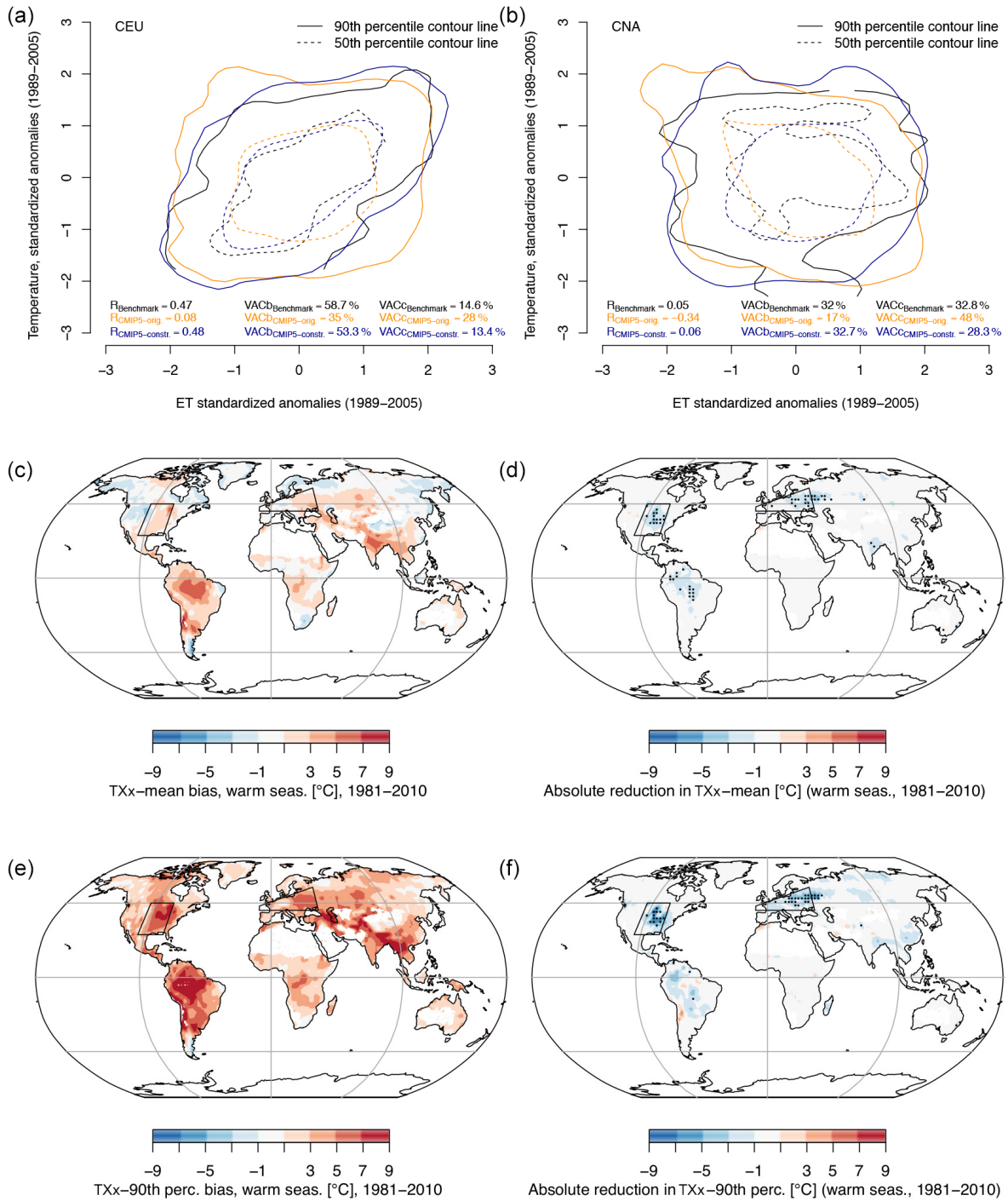
Figure 4. (a, b) Contour lines of bivariate kernel density estimates of T–ET relationship in the benchmarking datasets and the original and constraint CMIP5 ensemble for **(a)** central Europe and **(b)** central North America (1989–2005, area average). **(c, e)** Biases in warm season **(c)** TXx mean and **(e)** 90th percentile of TXx in the original CMIP5 ensemble and **(d, f)** reduction in **(d)** TXx mean and **(f)** 90th percentile TXx through the application of the land-coupling constraint. Regions with a significant reduction in **(d)** TXx mean and **(f)** the across-model average in the 90th percentile of TXx according to a permutation significance test are stippled.
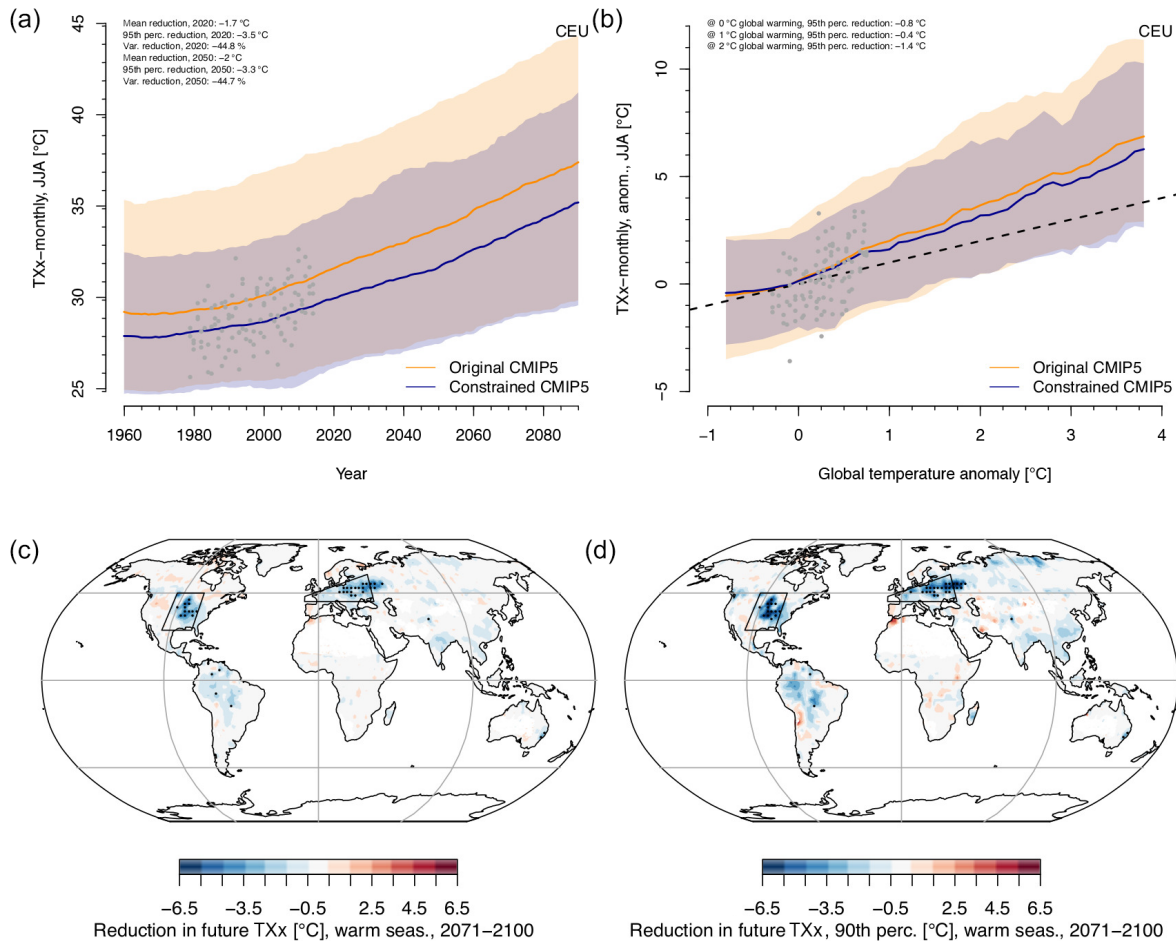
Figure 5. Application of land-coupling constraint to CMIP5 ensemble. **(a, b)** Ensemble prediction of original and constrained multi-model ensemble for **(a)** future absolute TXx and **(b)** range of TXx anomalies relative to global mean temperature anomalies in each model, following Seneviratne et al. (2016). Envelopes indicate 5th to 95th percentile. **(c, d)** Global maps of projected changes in simulated **(c)** mean TXx and **(d)** 90th percentile of TXx in the VACc-constrained CMIP5 ensemble.

3.2.2 Is there a link between present-day land–atmosphere coupling and warming projections?

We investigate whether the representation of land–atmosphere coupling in climate models affects the magnitude of 21st-century warming (e.g. Fischer et al., 2012; Stegehuis et al., 2013). We first note that regions sensitive to land–atmosphere coupling in the CMIP5 model ensemble also show relatively strong warming in daily-scale temperature extremes (TXx), for example central America or southern and central Europe (Fig. 6a and b). More importantly, however, models that produce frequent VAC$_c$ occurrences (water-limited regimes) tend to be associated with larger rates of warming in TXx, although it should be emphasized that this relationship is not simple or linear (Fig. 6c and d; see also Fischer et al., 2012). Conversely, this pattern reverses in boreal regions, where strongly energy-limited models (i.e. very few VAC$_c$ occurrences) tend to produce

larger warming. However, in boreal regions this apparent relationship likely stems from a spurious correlation with the individual models' background warming (i.e. warming in annual averages), as the correlation in fact disappears if the background warming is subtracted from summer warming (Fig. S10). In contrast, in midlatitude regions warm season warming that exceeds annual average warming remains confined to the warm season. A multi-model projection constrained by a plausible representation of land–atmosphere coupling reduces differences in TXx estimates in a future climate relative to the present in coupling-sensitive regions such as central Europe and central North America locally by around 0.5 to 1 °C, but this remains a regional effect (Fig. 6e and f). These results are consistent with earlier studies that used an ensemble of regional models over Europe that used the standard deviation of temperatures as a constraint (Fischer et al., 2012).
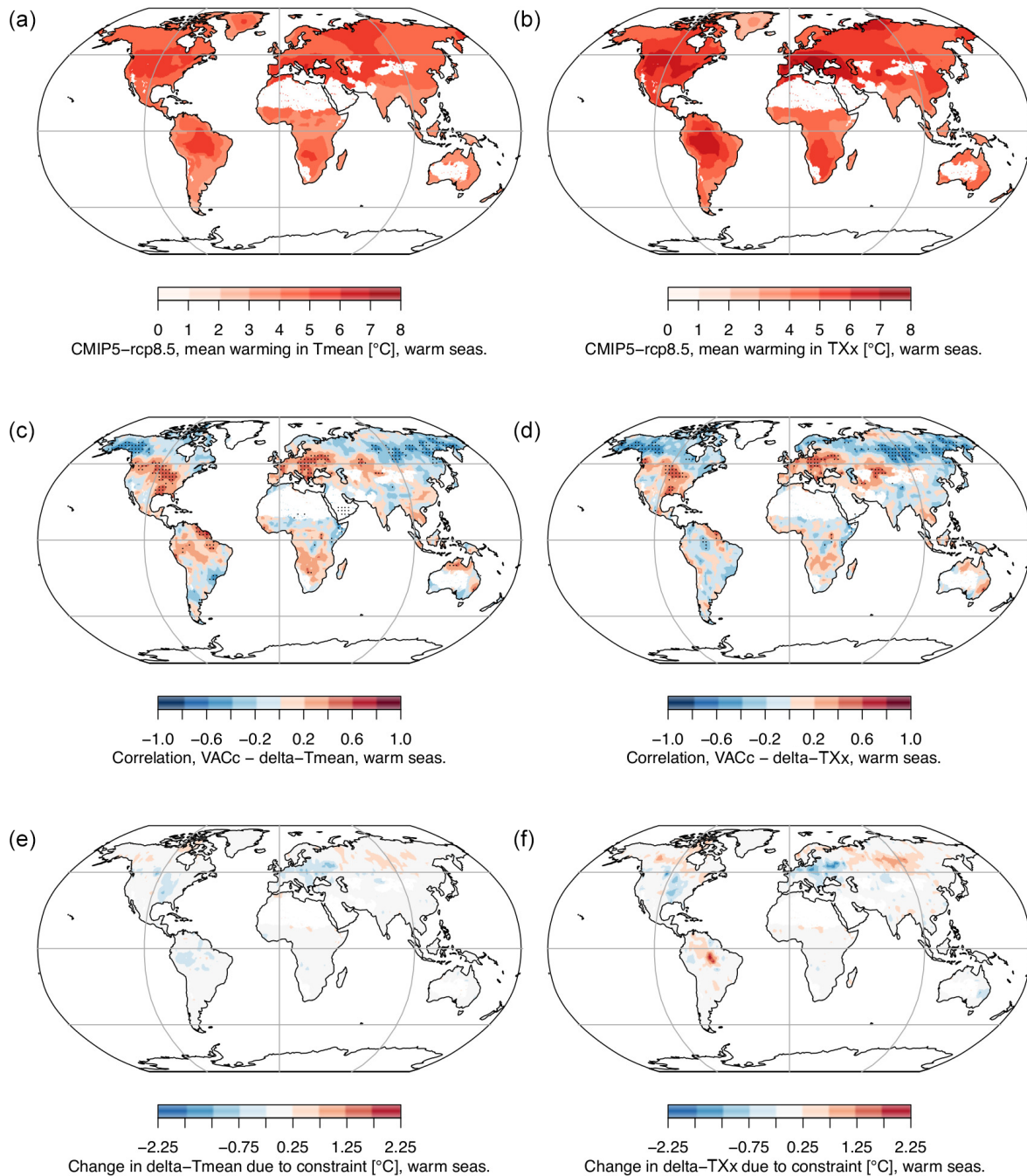
Figure 6. (a, b) Projected warming in warm season **(a)** mean temperature and **(b)** TXx across the CMIP5 ensemble (RCP8.5 scenario, 2071–2100 relative to 1981–2010). **(c, d)** Correlation between VACc in the warm season and the projected warming in **(c)** mean temperature and **(d)** TXx. Stippling indicates significant correlations. **(e, f)** Relative change in **(e)** mean warming and **(f)** TXx warming due to the application of the land–atmosphere coupling constraint; warming defined as 2071–2100 relative to 1981–2100.

4 Conclusions

In the present study, we have evaluated land–atmosphere coupling in state-of-the-art climate models with an ensemble of observations using a diagnostic based on coincidences of T and ET anomalies (the VAC index). While observa-tions and models broadly agree on spatial patterns of land–atmosphere coupling, our results reveal that models differ widely in coupling-sensitive regions in the midlatitudes and the tropics. Several models exhibit systematically too fre-quent coincidences of high temperature anomalies with neg-ative ET anomalies (water-limited regimes) in midlatitude

regions in the warm season and in several tropical regions year-round.

Across the multi-model ensemble, we found a strong association of land–atmosphere coupling with simulated temperature variability and extremes. The spread between models largely explains differences in simulated monthly temperature variability and daily extremes. We applied a land–atmosphere coupling constraint to the multi-model ensemble, which considerably improves the representation of land–atmosphere coupling in the ensemble and reduces biases in temperature variability and extremes in present-day simulations in a physically consistent manner (Fig. 4). Furthermore, the constraint leads to reduced variability and lower extreme temperatures in future projections. However, the overall projected changes in temperature extremes are not as strongly affected (reduction by around 0.5–1.0 °C locally in regions that are sensitive to land–atmosphere coupling) because the models with overestimated land–atmosphere coupling display similar anomalies from the multi-ensemble mean in the present and future. In conclusion, we selected models with a *physically plausible* representation of land surface processes (and near-surface climate) using observations-based constraints that are guided by physical considerations. This approach complements more traditional bias correction approaches and offers new avenues to obtain improved estimates of future climate impacts.

References

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth System Models, J Climate, 26, 6801–6843, 2013.

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

Boberg, F. and Christensen, J. H.: Overestimation of Mediterranean summer temperature projections due to model deficiencies, Nat. Clim. Change, 2, 433–436, 2012.

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

Bonan, G.: Ecological climatology: concepts and applications, Cambridge University Press, Cambridge, 2015.

Bosilovich, M. G.: NASA's modern era retrospective-analysis for research and applications: Integrating Earth observations, IEEE Earthzine, 1, 82367, 2008.

Cannon, A. J.: Multivariate Bias Correction of Climate Model Outputs: Matching Marginal Distributions and Inter-variable Dependence Structure, J. Climate, 29, 7045–7064, 2016.

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T.,

and Reichstein, M.: Global covariation of carbon turnover times with climate in terrestrial ecosystems, Nature, 514, 213–217, 2014.

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

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., and Luke, C. M.: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability, Nature, 494, 341–344, 2013.

Davin, E. L., Maisonnave, E., and Seneviratne, S. I.: Is land surface processes representation a possible weak link in current Regional Climate Models?, Environ. Res. Lett., 11, 074027, doi:10.1088/1748-9326/11/7/074027, 2016.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, 2011.

Dirmeyer, P. A.: The terrestrial segment of soil moisture–climate coupling, Geophys. Res. Lett., 38, L16702, doi:10.1029/2011GL048268, 2011.

Dirmeyer, P. A., Jin, Y., Singh, B., and Yan, X.: Trends in land–atmosphere interactions from CMIP5 simulations, J. Hydrometeorol., 14, 829–849, 2013.

Donges, J., Schleussner, C.-F., Siegmund, J., and Donner, R.: Event coincidence analysis for quantifying statistical interrelationships between event time series, Eur. Phys. J. Spec. Top., 225, 471–487, 2016.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Fischer, E. M. and Schär, C.: Future changes in daily summer temperature variability: driving processes and role for temperature extremes, Clim. Dynam., 33, 917–935, 2009.

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

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

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J. G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S. I., Walz, A., Wattenbach, M., Zavala, M. A., and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, Global Change Biol., 21, 2861–2880, 2015.

Grimm, V. and Railsback, S. F.: Pattern-oriented modelling: a "multi-scope" for predictive systems ecology, Philos. T. Roy. Soc. B, 367, 298–310, 2012.

Hall, A. and Qu, X.: Using the current seasonal cycle to constrain snow albedo feedback in future climate change, Geophys. Res. Lett., 33, L03502, doi:10.1029/2005GL025127, 2006.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, 2014.

Hauser, M., Orth, R., and Seneviratne, S. I.: Role of Soil Moisture vs. Recent Climate Change for the 2010 Heat Wave in Western Russia, Geophys. Res. Lett., 43, 2819–2826, doi:10.1002/2016GL068036, 2016.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dynam., 4, 219–236, doi:10.5194/esd-4-219-2013, 2013.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, Nat. Geosci., 4, 17–21, 2011.

IPCC: Summary for Policymakers, in: Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change, edited by: Field, C., Barros, V., Stocker, T., Dahe, Q., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G., Allen, S., Tignor, M., and Midgley, P., Cambridge University Press, Cambridge, 2012.

Jung, M., Reichstein, M., Margolis, H. A., et al.: Global patterns of land–atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res.-Biogeo., 116, G00J07, doi:10.1029/2010JG001566, 2011.

King, A. D., Black, M. T., Min, S.-K., Fischer, E. M., Mitchell, D. M., Harrington, L. J., and Perkins-Kirkpatrick, S. E.: Emergence of heat extremes attributable to anthropogenic influences, Geophys. Res. Lett., 43, 3438–3443, 2016.

Knutti, R.: Should we believe model predictions of future climate change?, Philos. T. Roy. Soc. Lond. A, 366, 4647–4664, 2008.

Knutti, R.: The end of model democracy?, Climatic Change, 102, 395–404, 2010.

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nat. Clim. Change, 3, 369–373, 2013.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194–1199, 2013.

Köppen, W.: Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt, Geogr. Z., 6, 593–611, 1900.

Koster, R. D., Dirmeyer, P. A., Guo, Z., et al.: Regions of strong coupling between soil moisture and precipitation, Science, 305, 1138–1140, 2004.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen–Geiger climate classification updated, Meteorol. Z., 15, 259–263, 2006.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the cou-

pled atmosphere–biosphere system, Global Biogeochem. Cy., 19, GB1015, doi:10.1029/2003GB002199, 2005.

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, Wiley Interdiscip Rev. Clim. Change, 5, 113–128, 2014.

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

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15, 967–981, doi:10.5194/hess-15-967-2011, 2011a.

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

Miralles, D. G., v. den Berg, M., Teuling, A., and d. Jeu, R.: Soil moisture–temperature coupling: A multiscale observational analysis, Geophys. Res. Lett., 39, doi:10.1029/2012GL053703, 2012.

Mitchell, D., Heaviside, C., Vardoulakis, S., Huntingford, C., Masato, G., Guillod, B. P., Frumhoff, P., Bowery, A., Wallom, D., and Allen, M.: Attributing human mortality during extreme heat waves to anthropogenic climate change, Environ. Res. Lett., 11, 074006, 2016a.

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., and Allen, M.: Realizing the impacts of a 1.5 °C warmer world, Nat. Clim. Change, 6, 735–737, 2016b.

Mueller, B. and Seneviratne, S.: Systematic land climate and evapotranspiration biases in CMIP5 simulations, Geophys. Res. Lett., 41, 128–134, 2014.

Mueller, B., Hirschi, M., and Seneviratne, S. I.: New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data, Hydrol. Process., 25, 996–1008, 2011a.

Mueller, B., Seneviratne, S., Jimenez, C., et al.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011b.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: 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, 2013.

Mystakidis, S., Davin, E. L., Gruber, N., and Seneviratne, S. I.: Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates, Global Change Biol., 22, 2198–2215, doi:10.1111/gcb.13217, 2016.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kato, K., Oyama, R., Ose, T., Kadokura, S., and Wada, K.: The JRA-25 reanalysis, J. Meteor. Soc. Jpn., 85, 369–432, 2007.

Orth, R., Zscheischler, J., and Seneviratne, S. I.: Record dry summer in 2015 challenges precipitation projections in Central Europe, Sci. Rep., 6, 28334, doi:10.1038/srep28334, 2016.

Otto, F. E., Haustein, K., Uhe, P., Coelho, C. A., Aravequia, J. A., Almeida, W., King, A., Coughlan de Perez, E., Wada, Y., Jan van Oldenborgh, G., Haarsma, R., Haustein, K., Uhe, P., v. Aalst, M., Aravequia, J. A., Almeida, W., and Cullen, H.: Factors other than climate change, main drivers of 2014/15 water shortage in southeast Brazil, B. Am. Meteorol. Soc., 96, S35–S40, 2015.

Perkins, S., Pitman, A., Holbrook, N., and McAneney, J.: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions, J. Climate, 20, 4356–4376, 2007.

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

Quiroga, R. Q., Kreuz, T., and Grassberger, P.: Event synchronization: a simple and fast method to measure synchronicity and time delay patterns, Phys. Rev. E, 66, 041904, doi:10.1103/PhysRevE.66.041904, 2002.

Rammig, A., Wiedermann, M., Donges, J. F., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M. D.: Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity, Biogeosciences, 12, 373–385, doi:10.5194/bg-12-373-2015, 2015.

Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., and Touré, A.: Assessment and enhancement of MERRA land surface hydrology estimates, J. Climate, 24, 6322–6338, 2011.

Reichstein, M., Mahecha, M. D., Ciais, P., Seneviratne, S. I., Blyth, E. M., Carvalhais, N., and Luo, Y.: Elk-testing climate–carbon cycle models: a case for pattern-oriented system analysis, iLEAPS Newslett., 11, 14–21, 2011.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, B. Am. Meteorol. Soc., 85, 381–394, 2004.

Rui, H. and Beaudoing, H.: README Document for Global Land Data Assimilation System Version 2 (GLDAS-2) Products, Goddard Earth Sciences (GES) Data and Information Services Center, https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/README_GLDAS2.pdf (last access: 29 May 2017), 2016.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., v. Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., v. d. Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, B. Am. Meteorol. Soc., 91, 1015, doi:10.1175/2010BAMS3001.1, 2010.

Santanello, J. A., Peters-Lidard, C. D., Kumar, S. V., Alonge, C., and Tao, W.-K.: Diagnosing the local land–atmosphere coupling in models and observations, iLEAPS Newslett., 9, 22–24, 2010.

Santanello, J. A., Roundy, J., and Dirmeyer, P. A.: Quantifying the land–atmosphere coupling behavior in modern reanalysis products over the US Southern Great Plains, J. Climate, 28, 5813–5829, 2015.

Schleussner, C.-F., Rogelj, J., Schaeffer, M., Lissner, T., Licker, R., Fischer, E. M., Knutti, R., Levermann, A., Frieler, K., and Hare, W.: Science and policy characteristics of the Paris Agreement temperature goal, Nat. Clim. Change, 6, 827–835, doi:10.1038/nclimate3096, 2016.

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

Seneviratne, S. I., Corti, T., Davin, E., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A.: Climate change and soil moisture-climate interactions: Using new diagnostics to identify hot spots of land–atmosphere coupling, iLEAPS Newslett., 9, 18–21, 2010a.

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

Seneviratne, S. I., Wilhelm, M., Stanelle, T., Hurk, B., Hagemann, S., Berg, A., Cheruy, F., Higgins, M. E., Meier, A., Brovkin, V., Claussen, M., Ducharne, A., Dufresne, J.-L., Findell, K. L., Ghattas, J., Lawrence, D. M., Malyshev, S., Rummukainen, M., and Smith, B.: Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment, Geophys. Res. Lett., 40, 5212–5217, 2013.

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

Sheather, S. J. and Jones, M. C.: A reliable data-based bandwidth selection method for kernel density estimation, J. Roy. Stat. Soc. B, 53, 683–690, 1991.

Sheffield, J. and Wood, E. F.: Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle, J. Geophys. Res.-Atmos., 112, D17115, doi:10.1029/2006JD008288, 2007.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, J Climate, 19, 3088–3111, 2006.

Sheffield, J., Wood, E. F., and Munoz-Arriola, F.: Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data, J. Hydrometeorol., 11, 253–275, 2010.

Siegmund, J. F., Wiedermann, M., Donges, J. F., and Donner, R. V.: Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species, Biogeosciences, 13, 5541–5555, doi:10.5194/bg-13-5541-2016, 2016.

Sillmann, J., Kharin, V., Zhang, X., Zwiers, F., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, J. Geophys. Res.-Atmos., 118, 1716–1733, 2013a.

Sillmann, J., Kharin, V., Zwiers, F., Zhang, X., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble:

Part 2. Future climate projections, J. Geophys. Res.-Atmos., 118, 2473–2493, 2013b.

Sippel, S., Lange, H., Mahecha, M., Hauhs, M., Bodesheim, P., Kaminski, T., Gans, F., and Rosso, O.: Diagnosing the Dynamics of Observed and Simulated Ecosystem Gross Primary Productivity with Time Causal Information Theory Quantifiers, PLoS One, 11, e0164960, doi:10.1371/journal.pone.0164960, 2016a.

Sippel, S., Otto, F., Forkel, M., Allen, M., Guillod, B., Heimann, M., Reichstein, M., Seneviratne, S., Thonicke, K., and Mahecha, M. D.: A novel bias correction methodology for climate impact simulations, Earth Syst. Dynam., 7, 71-88, doi:10.5194/esd-7-71-2016, 2016b.

Stegehuis, A. I., Teuling, A. J., Ciais, P., Vautard, R., and Jung, M.: Future European temperature change uncertainties reduced by using land heat flux observations, Geophys. Res. Lett., 40, 2242–2245, 2013.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M.: Climate Change 2013: The Physical Science Basis, Working Group 1 (WG1) Contribution to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5), Tech. rep., Cambridge, UK and New York, NY, 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. T. Roy. Soc. Lond. A, 365, 2053–2075, 2007.

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, Geosci. Model Dev., 9, 2809–2832, doi:10.5194/gmd-9-2809-2016, 2016.

Van Huijgevoort, M., Van Lanen, H., Teuling, A., and Uijlenhoet, R.: Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge, J. Hydrol., 512, 421–434, 2014.

Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P.: Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models, J. Geophys. Res.-Biogeo., 119, 794–807, 2014.

Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O., and Seneviratne, S. I.: Impact of soil moisture on extreme maximum temperatures in Europe, Weather Clim. Extr., 9, 57–67, 2015.

Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P.: Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05 spatial resolution, Water Resour. Res., 46, W05512, doi:10.1029/2009WR008716, 2010.

Zscheischler, J., Orth, R., and Seneviratne, S. I.: A submonthly database for detecting changes in vegetation-atmosphere coupling, Geophys. Res. Lett., 42, 9816–9824, 2015.