Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Open Access
Discussions

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 Zürich, 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 (e.g. water and carbon). 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 mid-latitude regions has been identified empirically for high-impact heatwaves, but individual climate models differ widely in their respective representation of land-atmosphere coupling. Here, we combine an ensemble of observations-based and simulated temperature (T) and evapotranspiration (ET) datasets and investigate coincidences of T anomalies with ET anomalies as a proxy for land-atmosphere interactions during periods of anomalously warm temperatures. We demonstrate that a relatively 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 mid-latitude regions during the warm season and in several tropical regions year-round. Further, we show that these coincidences (high T, low ET), as diagnosed by the land-coupling coincidence metrics, are closely related to the variability and extremes of simulated temperatures across a multi-model ensemble. Thus, our approach offers a physically consistent, diagnostic-based avenue to evaluate these ensembles, and subsequently reduce model biases in simulated and predicted extreme temperatures. Following this idea, we derive a land-coupling constraint based on the spread of 54 combinations of T-ET benchmarking datasets and consequently retain only a subset of CMIP5 models that produce a land-coupling behaviour that is compatible with these observations-based benchmark estimates. The constrained multi-model projections exhibit lower temperature extremes in regions where models show substantial spread in T-ET coupling, and in addition, biases in the climate model ensemble are consistently reduced.

1 Introduction

The exchange of matter and energy between the land surface and the atmosphere is a crucial feature of the Earth's climate (Bonan, 2015). 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 so-called "energy-limited" and "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,b), implying a positive association between near-surface temperature (T) and evapotranspiration (ET). In contrast, in a dry, water-limited state, the land controls near-surface climate through a lack of soil moisture, and a corresponding reduction in evapotranspiration and latent cooling (see conceptual Fig. 1a,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 mid-latitude 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; Lorenz et al., 2012; Dirmeyer et al., 2013; Seneviratne 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 at various spatial and temporal scales, which relate to

i. scarcity of accurate observational products of soil moisture or evapotranspiration at large spatiotemporal scales and relatively short observational periods (Mueller and Seneviratne, 2014),

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,b for a simple conceptual example), and further contribute to uncertainties related to projected increases in summer temperature variability in the 21st century in mid-latitude 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.,

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

2011; Lorenz et al., 2012). Conceptually, the notion of "diagnostic-based model evaluation" as discussed here is consistent with so-called "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 at 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 relevant, 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 criteria have been used to constrain carbon cycle projections (Cox et al., 2013; Wenzel et al., 2014; Mystakidis et al., 2016), to select models for event attribution analyses (Perkins et al., 2007; King et al., 2016; Otto et al., 2015), in the context of refining precipitation projections (Orth et al., 2016) or to resample large initial-condition ensembles to alleviate biases without distorting the multivariate structure of climate model output (Sippel et al., 2016b). However, 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 and subjective, and because good model performance w.r.t. any given metric does not translate directly into (more) reliable projections (Knutti, 2008).

Hence, the starting point for the present analysis, -in the sense of being necessary, but not sufficient to assure 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.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

2 Data & Methods

2.1 Datasets for T-ET coupling analysis and model evaluation

Global temperature and evapotranspiration datasets

In order to evaluate T-ET coupling in global climate models, an ensemble of 18 gridded evapotranspiration estimates, taken from the LandFlux-EVAL multi-data set 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 (Section 3). The ensemble of ET reference datasets has been generated by combining a wide range of different ET estimates, based on five diagnostic (observations-based) products, five land surface models driven by observations and four reanalysis products (Mueller et al., 2013). The three temperature datasets are based on one observational product (the Climate Research Unit dataset, (Harris et al., 2014)) and two reanalysis products (The ERA-Interim reanalysis, (Dee et al., 2011), and the National Center of Environmental Prediction (NCEP) reanalysis (Kalnay et al., 1996), see Table 1). As fewer temperature than evapotranspiration datasets are used for the present study, we have tested that the spread between individual temperature datasets is substantially smaller than the differences between individual ET products. Therefore, the 54 T-ET coincidence datasets (denoted as "T-ET coupling benchmarks" in the remainder of the paper) represent a relatively large spread of plausible T-ET coupling estimates, but it should be emphasized that the datasets are not independent realizations. Thus, we use the spread of this observations-based ensemble of T-ET datasets as a measure of uncertainty, 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 Section 3.2 we use ERA-Interim (Dee et al., 2011) as a reference dataset.

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 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). We have tested that individual ensemble members from the same model tend to show a comparably small spread in VAC-coupling, indicating that the large spread across models likely arises from differences in model structure.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Open Access
Discussions

Data processing and analysis

All datasets were remapped to a common 2.5°x2.5° spatial resolution for analysis and before computing T-ET coincidences. For model evaluation (Section 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 so-called IPCC-SREX regions (IPCC, 2012).

2.2 Diagnostic-based model evaluation using T-ET coupling

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, and the latter is often diagnosed by correlation-based metrics such as for example the Pearson correlation between T and ET, $\rho_{(T,ET)}$ (Seneviratne et al., 2006; Lorenz et al., 2012). Here, we aim to exploit the T-ET coupling by using a natural extension of $\rho_{(T,ET)}$ that focusses on the tails of T-ET dependedencies. Deseasonalized and detrended time series of ET (x_i^{ET}) and T (x_i^T, with i and N denoting the time step and time series length, respectively), 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}:

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

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

$$Pr[X > th_{upper}] = Pr[X \leq th_{lower}] = p \quad (1)$$

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 $Pr[X \leq th_{lower}] = Pr[X > th_{upper}] = 0.3$). Here, we focus on coincidences of *warm temperature anomalies* ("T-events": $x_i^T > th_{upper}^T$) with anomalies in ET ("ET-events", i.e. either $x_i^{ET} > th_{upper}^{ET}$ for VAC_b or $x_i^{ET} < th_{lower}^{ET}$ for VAC_c), i.e. we derive coincidence rates r_{VAC_b} by counting the number of VAC_b-events (see Quiroga et al., 2002;

Donges et al., 2016, for earlier formulations of event coincidence analysis):

$$r_{VAC_b} = \frac{1}{N_0}\sum_{i=1}^{N} 1_{[VAC_b]}(x_i^{VAC})$$

Here, $1_A(x)$ is the indicator function, defined as $1_A(x) = 1$ if $x \epsilon A$ and $1_A(x) = 0$ otherwise. N_0 acts as a normalization constant and is chosen in our study such that $0 \leq r_{VAC_b} \leq 1$, i.e. we normalize with the total number of "T-events", $N_0 = \sum_{i=1}^{N} 1_{[x^T > th^T_{upper}]}(x_i^T)$. In other words, if all (none) of the "T-events" in the time series would coincide with "ET-events" (defined for VAC_b), then the average coincidence rates would be given by $r_{VAC_b} = 1$ ($r_{VAC_b} = 0$). 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. $r_{VAC_b} = \frac{1}{N}\sum_{i=1}^{N} 1_{[x_i^{ET} > th^{ET}_{upper}]}(x_i^{ET})$. Coincidence rates r_{VAC_c} follow equivalently by replacing VAC_b with VAC_c and in the definition of "ET-events" in the previous description. We compute r_{VAC_b} and r_{VAC_c} for all seasons but with an emphasis on the warmest periods of the year. Fig. 1 shows a simple example of monthly time series of T and ET simulated from two CMIP5 models for the same location (area-averaged over Central Europe, CEU), and occurrences of VAC_b and VAC_c are highlighted. Please note that event coincidence analyses are frequently applied in the context of ecosystem science (e.g. Rammig et al., 2014; Siegmund et al., 2016).

In comparison to more traditional coupling metrics, such as e.g. $\rho_{(T,ET)}$, VAC_{a-d} might be expected to yield similar results on very long time scales, whereas on shorter time scales the VAC_{a-d} index might pick up non-linearities in the tails (e.g. during warm temperature anomalies). We note that on the monthly time scale (as used in the present study), distinct non-linearities are detected in models and observations in summer T-ET coupling e.g. in Central Europe, where a larger number of VAC_c events occurs than that would be inferred from a correlation-based metric $\rho_{(T,ET)}$ (Fig. S1). However, $\rho_{(T,ET)}$ yields qualitatively similar results. In addition to the main text, the model evaluation is presented for a 90th percentile threshold, and for $\rho_{(T,ET)}$ to demonstrate robustness to the chosen methodological approach (cf. Fig. S2).

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_{VACc}) 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 the range of observational uncertainty.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

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 (Section 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 (Section 3.2).

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

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 Central Europe through both r_{VAC_b} and r_{VAC_c} across various seasons (Fig. 2, top) or diagnosed through more traditional coupling metrics such as $\rho_{(T,ET)}$ (Appendix B). Individual models indicate pronounced qualitative differences in the warm season, where some models point to energy-limited, whereas others point to predominantly water-limited conditions (Fig. 2, top, and Fig. 1, for an illustrative example). Observations-based T-ET datasets agree qualitatively, i.e. indicating energy-limited to neutral conditions in the Central European example, thus implying an overestimation of water-limited regimes in Central Europe 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. 2, see Fig. S2 for a definition of the warm season in each pixel). In mid-latitude 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. These discrepancies hold also if metrics that emphasize the whole distribution ($\rho_{(T,ET)}$) or more extreme parts of the tail (VAC based on a 90th percentile threshold) are used for model evaluation (Figs. S3-S5). 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. 2).

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 tend to lie relatively close to each other (Figs. S6-S7). However, model-specific signatures of model output are not unusual, as diagnosed before e.g. 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. S6). A detailled 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 Fig. S6-S7. 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

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

water-limited regimes in individual models w.r.t. observations does not exclude the possibility of strong water limitations during extreme events in the real world (Miralles et al., 2012; Whan et al., 2015) or possible future changes of the coupling strength. Further, we note that observations-based benchmark datasets also 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).

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 of temperatures at monthly and inter-annual time scales across the CMIP5 multi-model ensemble in Central Europe (Fig. 3a) and in most regions of the globe except in some subarctic climates (Fig. 3b). Therefore, this relationship is indicative for 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 $VACc$ indices within the range of benchmark datasets also produce a realistic near surface temperature variability, whereas models that fall too frequently in water-limited regimes also overestimate summer temperature variability (Fig. 3a). Moreover, in mid-latitude and tropical regions, the state of the land surface is strongly associated with the mean and variability of temperature extremes at the daily time scale in the warmest season (TXx, Fig. 3c,d). The link between 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 constraint" (the likely root cause for 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

A constraint on land-atmosphere coupling in the CMIP5 ensemble.

The constrained ensemble resembles the observational datasets in land-atmosphere coupling (Fig. 4a-b, Appendix S9-S11 for details), and a corresponding improvement in the representation of temperature extremes at the daily time scale would be expected due to the intimate link between land-atmosphere coupling and temperature variability and extremes (see previous Section).

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. 4). 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 Central North America, Central Europe or the Amazon (relative to

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

ERA-Interim, see Fig. 4). In a CMIP5 ensemble constrained by the land-atmosphere coupling metric VAC_c, the representation of temperature extremes is substantially improved in regions prone to coupling-induced biases (Fig. 4). 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. 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.

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 for biases into account are still widely lacking (Ehret et al., 2012; Bellprat et al., 2013), the identification of models with a *physically plausible* representation of near-surface climate and land-atmosphere interactions at 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 motivated 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). Fig. S12 indicates that VAC_c occurrences across the CMIP5 ensemble are negatively associated with precipitation and evapotranspiration in the warm season in mid-latitude regions - 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 evapotranspiration or rainfall.

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 South and Central Europe (Fig. 6, top). 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 (middle panel in Fig. 6, Fig. S13, e.g. 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. S13). In contrast, in mid-latitude regions warm season warming that exceeds annual average warming remains confined to the warm season (Fig. S13). A multi-model projection constrained by a plausible representation of land-atmosphere coupling reduces future TXx estimates in coupling-sensitive regions such as Central Europe and Central North America by up to 1.5°C. 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).

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

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 large temperature and evapotranspiration anomalies. While observations and models broadly agree on spatial patterns of land-atmosphere coupling, our results reveal that models differ widely in coupling-sensitive regions in the mid-latitudes and the tropics. Several models exhibit systematically too frequent coincidences of high temperature anomalies with negative ET anomalies (water-limited regimes) in mid-latitude 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 reduces biases in temperature variability and extremes in present-day simulations in a physically consistent manner, and leads to reduced variability and lower extreme temperatures in future projections. 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.

Acknowledgements. We appreciate the creators and providers of all data sets, including the University of East Anglia Climate Research Unit (CRU) for providing temperature data (British Atmospheric Data Centre, 2008, available from http://badc.nerc.ac.uk/data/cru), ERA-Interim data provided courtesy ECMWF, and NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/). Furthermore, we acknowledge the Global Modeling and Assimilation Office and the GES DISC for the dissemination of MERRA and MERRA-LAND. The CFSR data are from the Research Data Archive which is maintained by the Computational and Information Systems Laboratory at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation. The original data are available from the RDA (http://dss.ucar.edu) in data set number ds093.0. We would further like to acknowledge the Japanese 25yr ReAnalysis and JMA Climate Data Assimilation System (JCDAS) for the dissemination of JRA-25 data. We thank the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This research has received funding from the European Commission Horizon 2020 BACI Project (Towards a Biosphere Atmosphere Change Index, grant No. 640176), and the European Research Council (ERC) under grant agreement No. 617518 (DROUGHT-HEAT). S.S. is grateful to the German National Academic Foundation for PhD funding.

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, Journal of 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, Geophysical Research Letters, 35, 2008.

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

Bosilovich, M. G.: NASA's modern era retrospective-analysis for research and applications: Integrating Earth observations, IEEE Earthzine, 1, 82 367, 2008.

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

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., et al.: 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, Geophysical Research Letters, 39, 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?, Environmental Research Letters, 11, 074 027, http://stacks.iop.org/1748-9326/11/i=7/a=074027, 2016.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, 2011.

Dirmeyer, P. A.: The terrestrial segment of soil moisture–climate coupling, Geophysical Research Letters, 38, 2011.

Dirmeyer, P. A., Jin, Y., Singh, B., and Yan, X.: Trends in land–atmosphere interactions from CMIP5 simulations, Journal of Hydrometeorology, 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, The European Physical Journal Special Topics, 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, 2012.

Fischer, E., Seneviratne, S., Lüthi, D., and Schär, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, Geophysical Research Letters, 34, 2007.

Fischer, E., Rajczak, J., and Schär, C.: Changes in European summer temperature variability revisited, Geophysical Research Letters, 39, 2012.

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

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Velde, M., Vicca, S., Babst, F., et al.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, Glob. Chang. Biol., 21, 2861–2880, 2015.

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

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, Geophysical Research Letters, 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, 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, Nature Geoscience, 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, 2012.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., 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., G: Biogeosci., 116, 2011.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–471, 1996.

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, Geophysical Research Letters, 43, 3438–3443, 2016.

Knutti, R.: Should we believe model predictions of future climate change?, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 366, 4647–4664, 2008.

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

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

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

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

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C., Kanae, S., Kowalczyk, E., Lawrence, D., 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, Meteorologische Zeitschrift, 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 coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, 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, Journal of Geophysical Research: Atmospheres, 117, 2012.

Miralles, D., De Jeu, R., Gash, J., Holmes, T., and Dolman, A.: Magnitude and variability of land evaporation and its componentsat the global scale, Hydrology and Earth System Sciences, 2011a.

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, 2011b.

Miralles, D., den Berg, M. v., Teuling, A., and Jeu, R. d.: Soil moisture-temperature coupling: A multiscale observational analysis, Geophysical Research Letters, 39, 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, Environmental Research Letters, 11, 074 006, 2016a.

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., and Allen, M.: Realizing the impacts of a 1.5 [deg] C warmer world, Nature Climate Change, 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, Hydrological Processes, 25, 996–1008, 2011a.

Mueller, B., Seneviratne, S., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J., Guo, Z., et al.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophysical Research Letters, 38, 2011b.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P., Dolman, A., Fisher, J., Jung, M., Ludwig, F., Maignan, F., et al.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrology and Earth System Sciences, 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 biology, 2016.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., et al.: The JRA-25 reanalysis, J. Meteorol. 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, Scientific Reports, 6, 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., et al.: Factors other than climate change, main drivers of 2014/15 water shortage in southeast Brazil, Bulletin of the American Meteorological Society, 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, Journal of 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, Nature Climate 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, Physical review E, 66, 041 904, http://journals.aps.org/pre/pdf/10.1103/PhysRevE.66.041904, 2002.

Rammig, A., Wiedermann, M., Donges, J., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M.: Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity, Biogeosciences, 12, 373–385, 2014.

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, Journal of 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 Newsletter, 11, 14–21, 2011.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015, 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 Newsletter, 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, Journal of 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, Nature Climate Change, 2016.

Seneviratne, S., 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 Newsletter, 9, 18–21, 2010a.

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. 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-Science Reviews, 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., et al.: Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment, Geophysical Research Letters, 40, 5212–5217, 2013.

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., and Wilby, R. L.: Allowable CO2 emissions based on regional and impact-related climate targets, Nature, 2016.

Sheather, S. J. and Jones, M. C.: A reliable data-based bandwidth selection method for kernel density estimation, J. R. Stat. Soc. Series B Stat. Methodol., 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, Journal of Geophysical Research: Atmospheres, 112, 2007.

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, 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, Journal of Geophysical Research: Atmospheres, 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, doi:doi:10.1371/journal.pone.0164960, accepted, 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 System Dynamics, 7, 71, 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, Geophysical Research Letters, 40, 2242–2245, 2013.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., et al.: Climate Change 2013: The Physical Science Basis. Working Group 1 (WG1) Contribution to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5). Cambridge, United Kingdom and New York, NY, Tech. rep., 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485, 2012.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 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, Geoscientific Model Development, 9, 2809–2832, doi:10.5194/gmd-9-2809-2016, http://www.geosci-model-dev.net/9/2809/2016/, 2016.

Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P.: Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models, Journal of Geophysical Research: Biogeosciences, 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 and Climate Extremes, 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 Resources Research, 46, 2010.

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

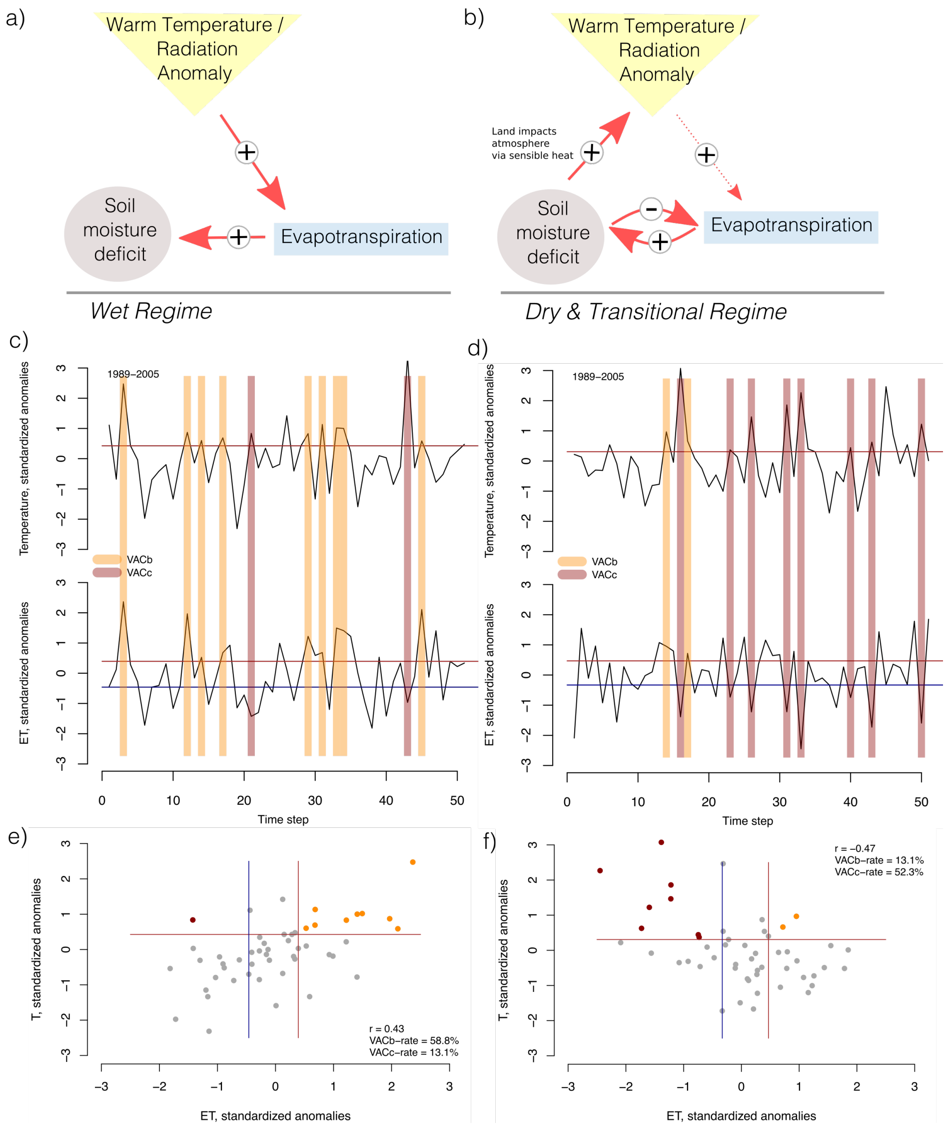

Figure 1. Conceptual figure of contrasting warm season temperature-evapotranspiration (T-ET) coupling in global climate models. (a, b) T-ET coupling in (a) dry & transitional regions and (b) wet regions, where soil moisture-temperature interactions play contrasting roles. (c-f) Contrasting T-ET coupling behaviour in a coupling-sensitive mid-latitude region in summer (Central Europe, spatial average, JJA, 1989-2005) in two different CMIP5 models (left, predominantly wet regime: NorESM1-M; right, predominantly dry regime: ACCESS1-3), illustrated as time series (c-d) and in the T-ET plane (e-f). Red lines in (c-f) indicate th_{upper} for T and ET, blue lines indicate th^{T}_{lower} (70th and 30th percentile in each individual time series, respectively).

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

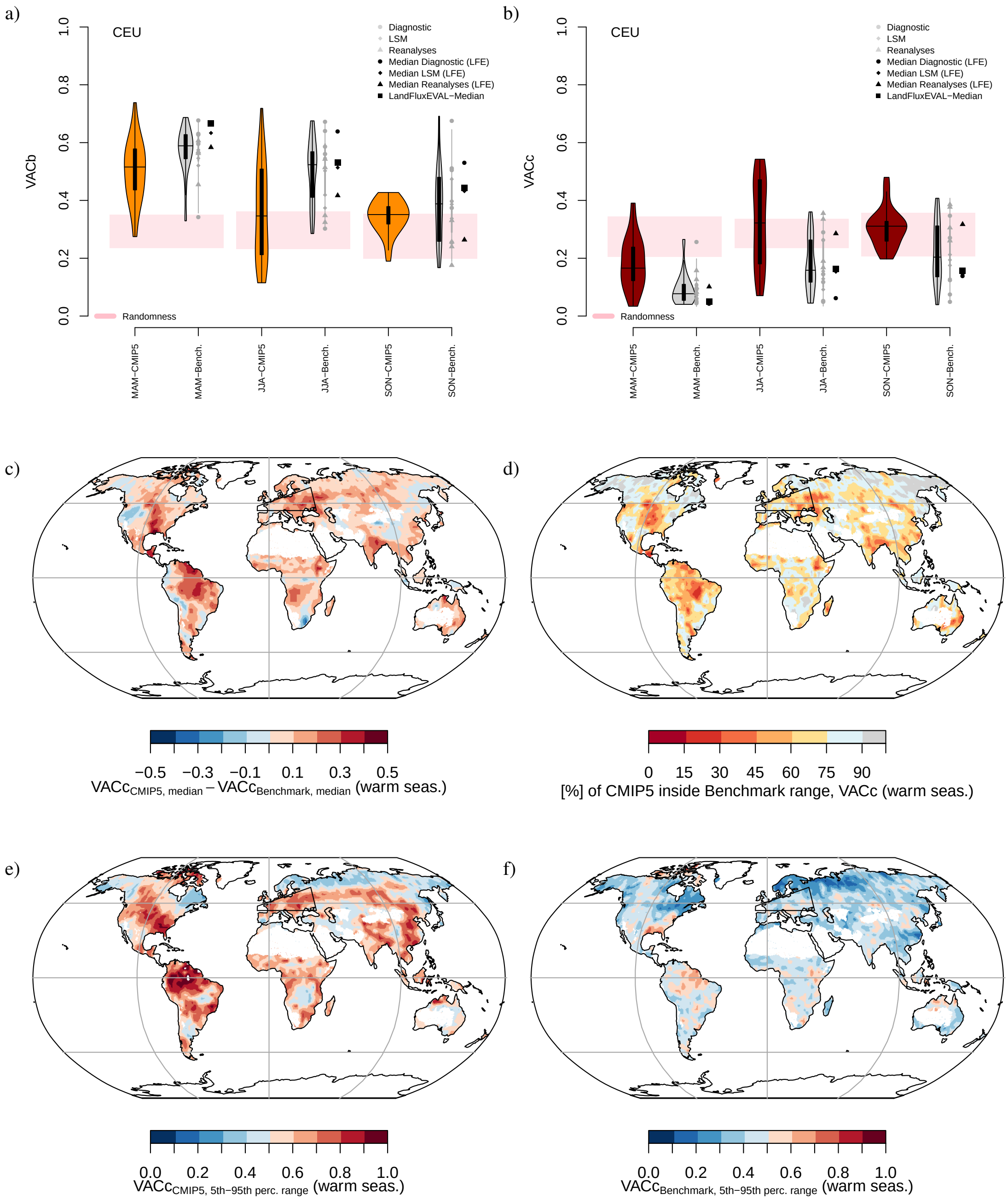

Figure 2. Evaluation of T-ET coupling in global climate models. (a, b) VACb (a) and VACc coupling in the CMIP5 climate model ensemble and observations-based benchmarking datasets in Central Europe (CEU, 1989-2005, area-average) with systematic warm season differences. Randomness indicates the 5th to 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).

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Discussions
Open Access

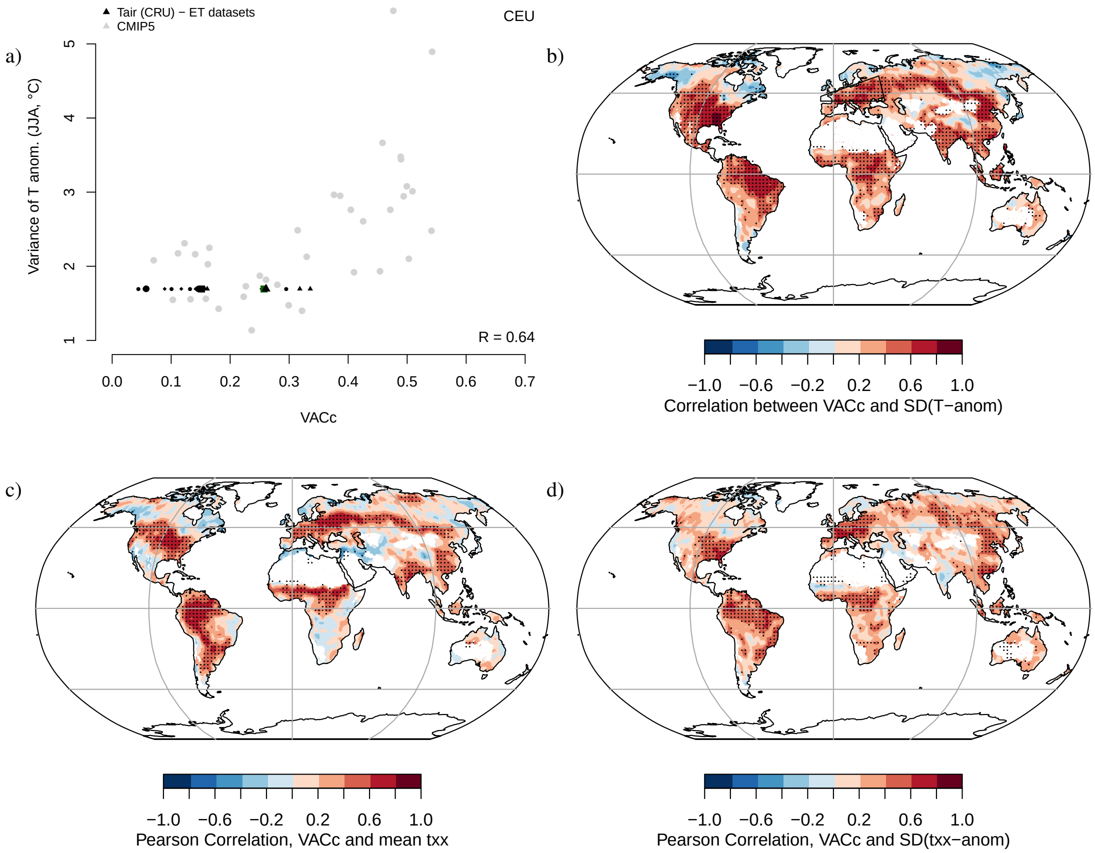

Figure 3. (a, b) Relationship between model-specific T-ET coupling (expressed through VACc) and model simulated variability of monthly temperature anomalies (JJA) in Central Europe (a), and globally (b). (c, d) Relationship betweeen VACc-coupling and mean (c) and standard deviation (d) of simulated monthly maximum value of daily maximum temperature (TXx) in summer (JJA).

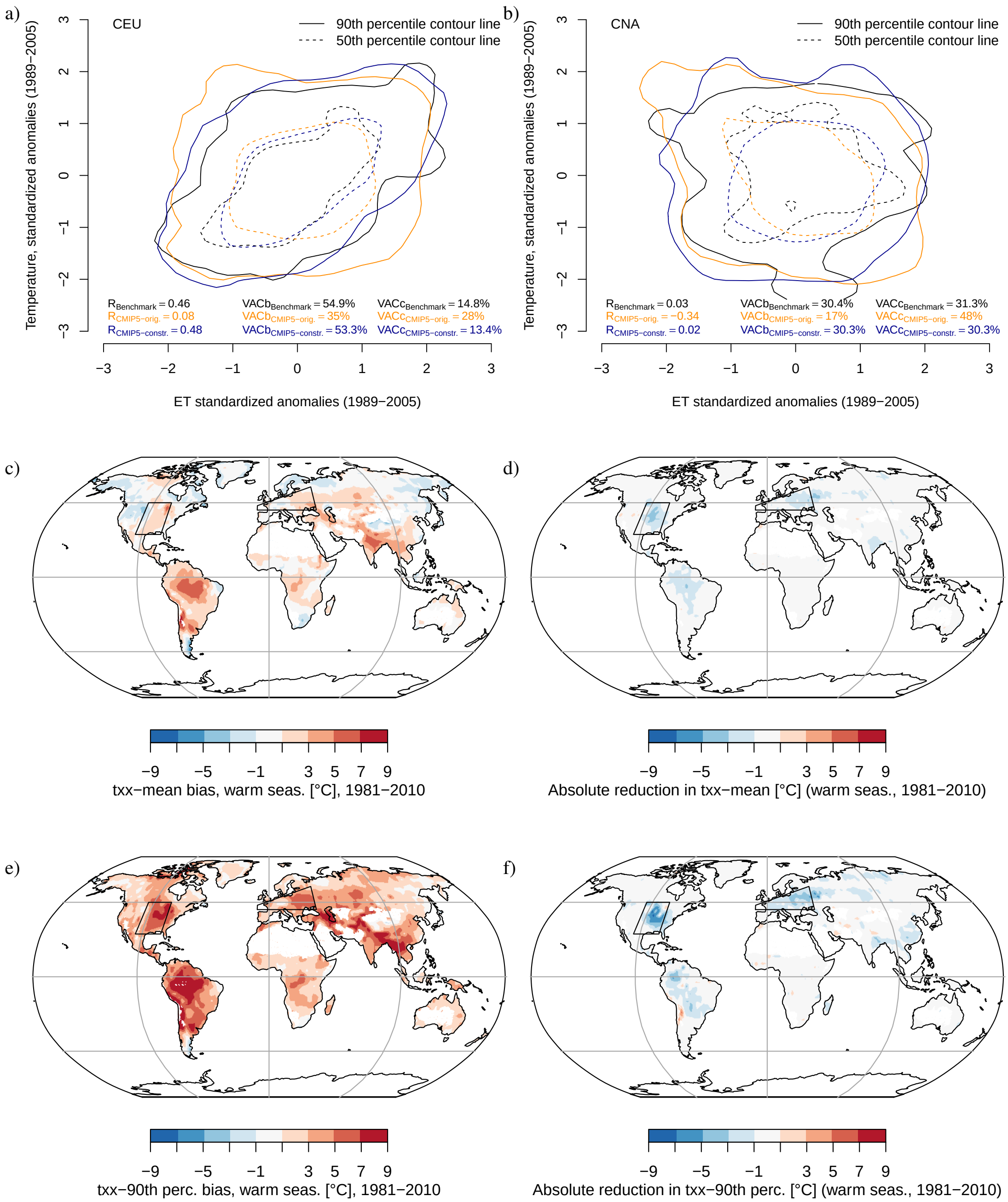

Figure 4. (a-b) Contour lines of bivariate kernel density estimates of T-ET relationship in the benchmarking datasets, the original and constraint CMIP5 ensemble for (a) Central Europe, and (b) Central North America (1989-2005, area-average). (c, e) Biases in warm season (c) TXx mean, and (e) 90th percentile of TXx in the original CMIP5 ensemble, and (d, f) reduction in biases in (d) TXx mean, and (f) 90th percentile TXx through the application of the land-coupling constraint.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Earth System Dynamics
Discussions

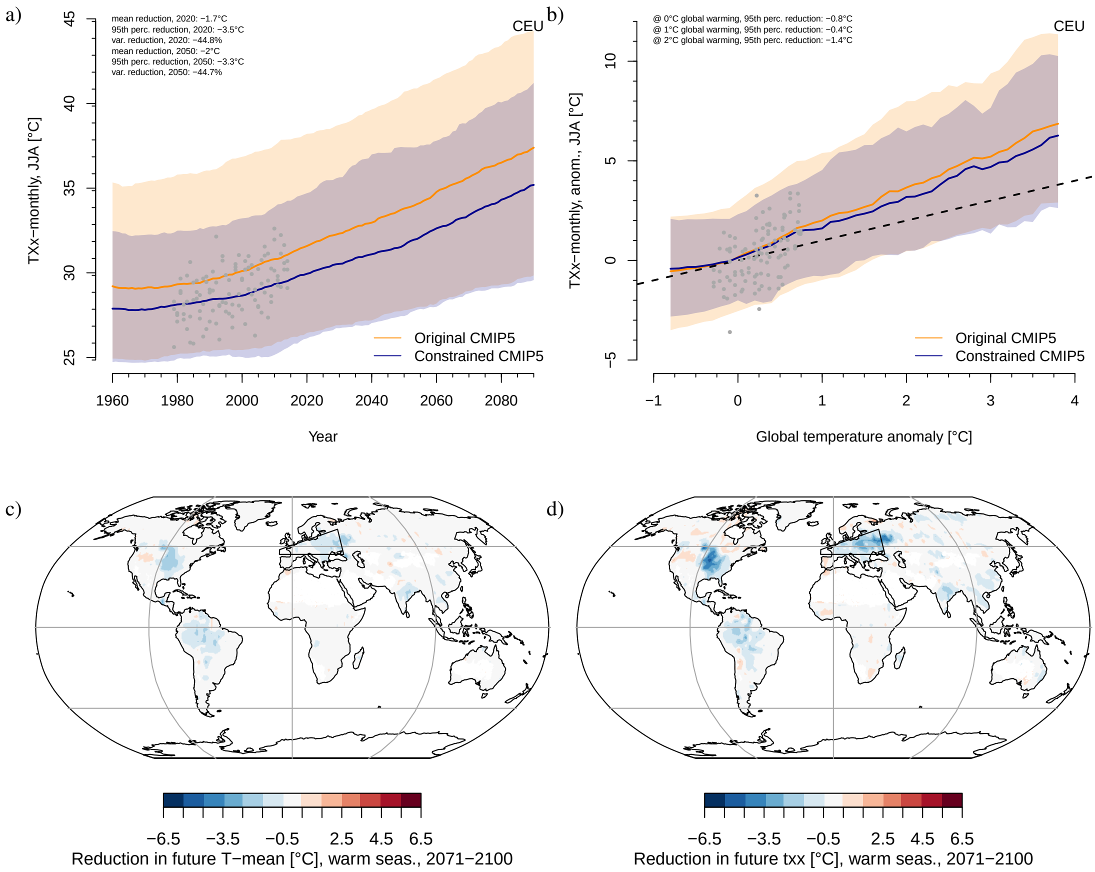

Figure 5. Application of land coupling constraint to CMIP5 ensemble. (a, b) Ensemble prediction of original and constrained multi-model ensemble for future absolute TXx (a) and range of TXx anomalies relative to global mean temperature anomalies (b), following Seneviratne et al (2016). Envelopes indicate 5th to 95th percentile. (c, d) Global maps of present-day and future changes in the simulated TXx 90th percentile in the VACc-constrained CMIP5 ensemble.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

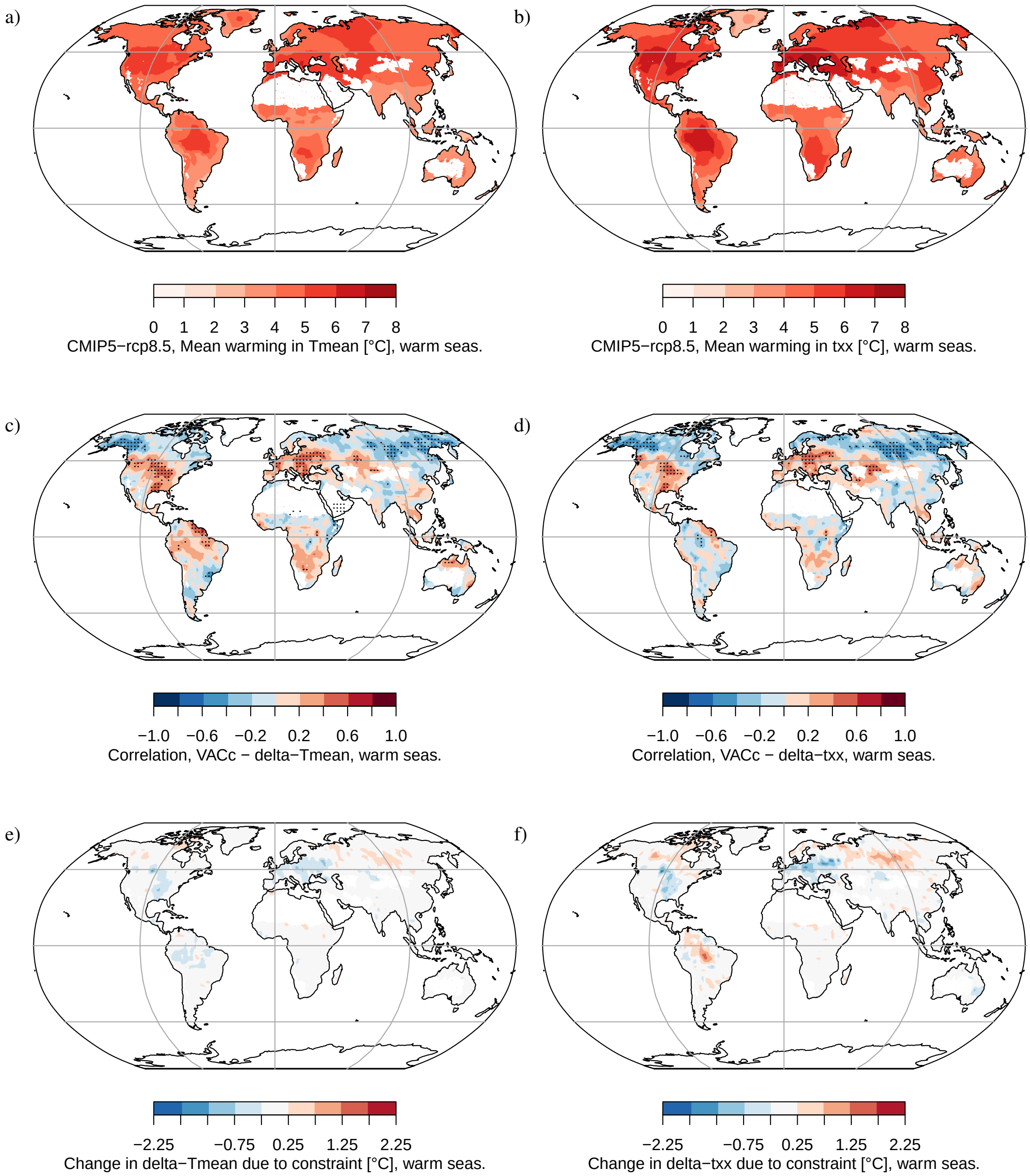

Figure 6. (a, b) Projected warming in warm season (a) mean temperature, and (b) TXx across the CMIP5 ensemble (RCP8.5 scenario, 2071-2100 relative to 1981-2010). (c, d) Correlation between VACc in the warm season and the projected warming in (c) mean temperature, and (d) TXx. (e, f) Relative change in (e) mean warming and (f) TXx warming due to the application of the land-atmosphere coupling constraint.

Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-48, 2016
Manuscript under review for journal Earth Syst. Dynam.
Published: 27 October 2016

Table 1. Datasets used for model evaluation

Name of dataset	Variable	Type / Group	Provider & 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)
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 and Wood (2007)
NOAH-PF[a,b]	ET	LSM	
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)
NCEP/DOE reanalysis 2[a]	T	Reanalysis	Kalnay et al. (1996)

[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).