# Evapotranspiration simulations in ISIMIP2a—Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets

Environmental Research Letters

LETTER

# Evapotranspiration simulations in ISIMIP2a—Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets

Richard Wartenburger[1,37], Sonia I Seneviratne[1], Martin Hirschi[1], Jinfeng Chang[24,30], Philippe Ciais[24], Delphine Deryng[2,3], Joshua Elliott[26], Christian Folberth[9], Simon N Gosling[20], Lukas Gudmundsson[1], Alexandra-Jane Henrot[14], Thomas Hickler[29,25], Akihiko Ito[23], Nikolay Khabarov[4], Hyungjun Kim[22], Guoyong Leng[8], Junguo Liu[12,4], Xingcai Liu[7], Yoshimitsu Masaki[18], Catherine Morfopoulos[28], Christoph Müller[17], Hannes Müller Schmied[5,6], Kazuya Nishina[16], Rene Orth[31,34], Yadu Pokhrel[13], Thomas A M Pugh[10,11], Yusuke Satoh[4], Sibyll Schaphoff[17], Erwin Schmid[19], Justin Sheffield[32,33], Tobias Stacke[15], Joerg Steinkamp[36], Qiuhong Tang[7], Wim Thiery[1,35], Yoshihide Wada[4], Xuhui Wang[24], Graham P Weedon[21], Hong Yang[27] and Tian Zhou[8]

1 Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, CH-8092 Zurich, Switzerland
2 Climate Analytics, 10969 Berlin, Germany
3 Columbia University Center for Climate Systems Research, New York, NY 10025, United States of America
4 International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
5 Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
6 Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, Germany
7 Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China
8 Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99352, United States of America
9 Ecosystem Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
10 School of Geography, Earth & Environmental Sciences and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, United Kingdom
11 Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Garmisch-Partenkirchen, Germany
12 School of Environmental Science and Engineering, South University of Science and Technology of China, Shenzhen, People's Republic of China
13 Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824 United States of America
14 Unité de Modèlisation du climat et des Cycles Biogéochimiques, UR SPHERES, Université de Liège, Quartier Agora, Liège, Belgium
15 Max Planck Institute for Meteorology, Hamburg, Germany
16 National Institute for Environmental Studies, Tsukuba, Japan
17 Potsdam Institute for Climate Impact Research (PIK), Telegraphenberg A31, 14473 Potsdam, Germany
18 Hirosaki University, Aomori, Japan
19 University of Natural Resources and Life Sciences, Department of Economics and Social Sciences, Feistmantelstrasse 4, A-1180 Vienna, Austria
20 School of Geography, University of Nottingham, Nottingham NG7 2RD, United Kingdom
21 Met Office (JCHMR), Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, United Kingdom
22 Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
23 National Institute for Environmental Studies, Tsukuba, Japan
24 Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ, Gif-sur-Yvette, France
25 Institute of Physical Geography, Geosciences, Goehte University, Frankfurt am Main, Germany
26 The University of Chicago, 5757 S. University Avenue, Chicago IL 60637, United States of America
27 Department of Systems Analysis, Integrated Assessment and Modelling, Eawag, 8600 Dübendorf, Switzerland
28 College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom
29 Senckenberg Biodiversity and Climate Research Centre (BiK-F) & Goethe-University Frankfurt, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany
30 Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, LOCEAN/IPSL, Paris, France
31 Department of Physical Geography, Bolin Centre for Climate Research, Stockholm University, SE-10691 Stockholm, Sweden
32 Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, United States of America
33 Geography and Environment, University of Southampton, Southampton, United Kingdom
34 Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, D-07745 Jena, Germany
35 Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
36 Zentrum für Datenverarbeitung, Johannes Gutenberg-Universität Mainz, Germany
37 Author to whom any correspondence should be addressed.

**E-mail:** richard.wartenburger@env.ethz.ch

**Keywords:** ISIMIP2a, evapotranspiration, uncertainty, cluster analysis, hydrological extreme events

OPEN ACCESS

RECEIVED
6 October 2017

REVISED
25 January 2018

ACCEPTED FOR PUBLICATION
15 May 2018

PUBLISHED
21 June 2018



## Abstract

Actual land evapotranspiration (ET) is a key component of the global hydrological cycle and an essential variable determining the evolution of hydrological extreme events under different

climate change scenarios. However, recently available ET products show persistent uncertainties that are impeding a precise attribution of human-induced climate change. Here, we aim at comparing a range of independent global monthly land ET estimates with historical model simulations from the global water, agriculture, and biomes sectors participating in the second phase of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2a). Among the independent estimates, we use the EartH2Observe Tier-1 dataset (E2O), two commonly used reanalyses, a pre-compiled ensemble product (LandFlux-EVAL), and an updated collection of recently published datasets that algorithmically derive ET from observations or observations-based estimates (diagnostic datasets). A cluster analysis is applied in order to identify spatio-temporal differences among all datasets and to thus identify factors that dominate overall uncertainties. The clustering is controlled by several factors including the model choice, the meteorological forcing used to drive the assessed models, the data category (models participating in the different sectors of ISIMIP2a, E2O models, diagnostic estimates, reanalysis-based estimates or composite products), the ET scheme, and the number of soil layers in the models. By using these factors to explain spatial and spatio-temporal variabilities in ET, we find that the model choice mostly dominates (24%–40% of variance explained), except for spatio-temporal patterns of total ET, where the forcing explains the largest fraction of the variance (29%). The most dominant clusters of datasets are further compared with individual diagnostic and reanalysis-based estimates to assess their representation of selected heat waves and droughts in the Great Plains, Central Europe and western Russia. Although most of the ET estimates capture these extreme events, the generally large spread among the entire ensemble indicates substantial uncertainties.

## 1. Introduction

Climate impact models are frequently used to quantify and analyse the effects of environmental changes in various socio-economic and environmental sectors under a given scenario design. However, the interpretative power of individual impact model simulations is severely limited due to the lack of thorough estimates of the full range of inter-model and inter-sectoral uncertainties Frieler *et al* (2015). The second phase of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2a) provides a new framework deemed to gain better uncertainty estimates to model-based projections through an integrative approach Warszawski *et al* (2014). For the critical assessment of extreme events, it is absolutely necessary to be aware of these uncertainties, concerning both the spread among the ISIMIP simulations as well as biases of the multi-model mean with respect to independent observation-based estimates from the recent past.

Key impact variables such as irrigation water demand or agricultural productivity are physically controlled by the partitioning of energy at the land surface, which largely depends on total evapotranspiration (ET, e.g. Betts *et al* 1996). As ET accounts for more than half of the precipitation fluxes in many regions, it is an important parameter controlling hydrological extreme events, in particular when considering its potential to amplify droughts and heat waves through coupling with soil moisture (e.g. Seneviratne *et al* 2010). However, to analyse such extreme events in greater detail, it is absolutely necessary to be aware of the full range of uncertainties inherent in different estimates of ET. In an early but comprehensive comparison of various land surface models within the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS, Henderson-Sellers *et al* 1993, Henderson-Sellers *et al* 1996), enormous uncertainties in the representation of evaporation among different land surface schemes have been found (Shao and Henderson-Sellers 1996, Chen *et al* 1997, Wood *et al* 1998, Pitman *et al* 1999). Although considerable progress has been made since Pitman (2003), land evaporation remains to be one of the most uncertain components of the global hydrological cycle to date (e.g. Fisher *et al* 2017). It is hence not surprising that a recent ensemble product of global ET estimates still reveals substantial uncertainties, which are comparable to the magnitude of uncertainties in precipitation estimates Mueller *et al* (2013). To thoroughly analyse extreme events within the ISIMIP2a framework, it is thus a prerequisite to precisely assess the magnitude of common ensemble statistics (mean, median and interquartile ranges, IQRs) of presently available ET estimates across datasets/models and sectors, and to further attempt to identify potential causes for differences between these estimates.

Uncertainties in estimates of ET can be due to multiple interrelated factors, including (but not limited to) the choice of the model or the forcing, the data category (model-based estimates from a specific sector, diagnostic or reanalysis-based estimates), the number of soil layers in the model and the ET scheme. Studies from the recent past have mainly focussed on the latter issue: the difficulties of choosing the most appropriate ET scheme for specific applications (e.g. Kay *et al* 2013, Bormann 2011), which can even translate into ambiguities in the interpretation of the evolution of past global drought conditions (Sheffield *et al* 2012,

Seneviratne 2012). Numerous papers confirm that the most widely used FAO56 Penman-Monteith formulation produces the most reasonable estimation of potential evaporation (Ortega-Farias *et al* 2004, DehghaniSanij *et al* 2004, Donohue *et al* 2010, Prudhomme and Williamson 2013), whereas others see similar performance of the Priestley-Taylor (Shaw and Riha 2011, Kingston *et al* 2009) or the Hargreaves formulation Kingston *et al* (2009), and a few do not make any clear recommendation (Schulz *et al* 1998, Kite and Droogers 2000). Other studies even state that Penman-Monteith does not essentially yield the best estimates (Weiß and Menzel 2008, Douglas *et al* 2009 and Xu and Chen 2005) suggest Priestley-Taylor, Vörösmarty *et al* (1998) suggest the Hamon method, and (Liu *et al* 2007, Liu and Yang 2010 and Sperna Weiland *et al* 2012) recommend the Hargreaves equation, where the latter suggests to use a re-calibrated form of the equation for climate change studies). Given these uncertainties, it is highly appropriate to assess in more detail whether the differences in ET schemes govern the overall differences across available ET estimates, or whether some of the other factors prevail.

Here we make use of the ISIMIP2a framework to assess selected hydrological extreme events by taking uncertainties into account that arise from difficulties in simulating global land ET. The analysis makes use of a large ensemble of ET datasets including models participating in the the global water, biomes and agriculture sectors, and also including other (independent) estimates from the following sources (introduced in more detail in section 2):

i. An ensemble of diagnostic datasets,
ii. the EartH2Observe Tier-1 dataset (version 1, Schellekens *et al* 2017, hereafter referred to as E2O),
iii. the LandFlux-EVAL initiative (Mueller *et al* 2013, hereafter abbreviated as LFE), and
iv. two recent global reanalyses that assimilate land surface variables or utilize time integrations of surface meteorological conditions (hereafter referred to as land reanalyses): ERA-Interim/Land Balsamo *et al* (2015) and MERRA-2 Reichle *et al* (2017).

This variety of datasets allows us to assess the full range of uncertainties captured by simulated or observation-based historical records of ET, and to identify strengths and weaknesses of individual groups of datasets with respect to their potential to reflect regional ET anomalies during extreme events. The ensemble is particularly suitable to draw valuable conclusions on structural differences among the datasets (e.g. to measure the influence of the meteorological forcing dataset employed in each of the model-based estimates).

The remainder of this paper is structured as follows. After a brief introduction to all input datasets in section 2, we present our methodological approach (section 3). This is followed by the presentation and discussion of spatial and spatio-temporal variabilities across the analysed datasets by means of a cluster analysis (section 4.1), after which we examine individual time series of global and regional ET (section 4.2). Section 5 summarizes the main findings of this study.

## 2. Data

Table 1 lists all datasets used in this study. In total, we assess 11 diagnostic datasets, nine models from E2O, 12 models from the ISIMIP2a agriculture sector, seven models from the ISIMIP2a biomes sector, 11 models from the ISIMIP2a global water sector, two land reanalyses and one composite dataset (the latter consisting of four different realizations). Note that ET estimates from all datasets except from the ISIMIP2a crop model simulations (labelled with asterisks (∗) in table 1) correspond to estimates of total ET at monthly resolution (hereafter denoted as $ET_{tot}$). Besides other information, the table also shows the category that each dataset is associated with. In the following, we provide some more information for each of the data categories. Please refer to the publications listed in table 1 for more detailed information on individual datasets.

### 2.1. Model simulations and land reanalyses

We analyse ET from historical simulations of the ISIMIP2a project using all models from the global water, biomes and agriculture sectors that simulate ET using the model's default ET scheme. The simulations used in this analysis are based on three distinct meteorological forcing datasets: the Global Soil Wetness Project Phase 3 (GSWP3, http://hydro.iis.u-tokyo.ac.jp/GSWP3/), the Princeton Global Meteorological Forcing Dataset version 2 (PGMFD v.2, Sheffield *et al* 2006) and the Water and Global Change (WATCH, Weedon *et al* 2011) Forcing Data methodology applied to ERA-Interim data (WFDEI, Weedon *et al* 2014). We use naturalized runs (i.e. without human impact) only, except from a few simulations from biomes models (see table 1). Due to the large number of models involved, we cannot provide a detailed description of each model here, and rather suggest the interested reader to either consider the publications listed in table 1 or to consult the summary information listing the most relevant characteristics of each model (freely accessible via the ISIMIP website, www.isimip.org/impactmodels/).

From the pool of crop model simulations of the ISIMIP2a agriculture sector, we only select harmonized simulations (see Elliott *et al* 2015 for details). All crop model simulations use fixed fertilizer application rates except for LPJmL and LPJ-GUESS, as fertilizer input is irrelevant for this version of these models. As previously mentioned

**Table 1.** Overview of datasets used in this study sorted by data category and dataset name. ISIMIP2a simulations are printed in bold.

| | | Meteorological forcing | | | |
|---|---|---|---|---|---|
| dataset | Data category | GSWP3 | PGMFD | WFDEI | Citation |
| FLUXNET-MTE | Diagnostic | | | | Jung *et al* (2009) |
| GETA 2.0 | Diagnostic | | | | Ambrose and Sterling (2014) |
| GLEAM V3.1A | Diagnostic | | | | Martens *et al* (2016) |
| MODIS Global ET | Diagnostic | | | | Mu *et al* (2011) |
| PM-MU-LANDFLUX | Diagnostic | | | | Vinukollu *et al* (2011) |
| PML-CSIRO | Diagnostic | | | | Zhang *et al* (2016) |
| PT-FI-LANDFLUX | Diagnostic | | | | Fisher *et al* (2008) |
| SEBS-LANDFLUX | Diagnostic | | | | Su (2001) |
| WANG-ET | Diagnostic | | | | Wang *et al* (2010a, b) |
| WB-MTE | Diagnostic | | | | Zeng *et al* (2014) |
| WECANN | Diagnostic | | | | Alemohammad *et al* (2016) |
| HBV-SIMREG | E2O | | | x | Beck *et al* (2016) |
| HTESSEL | E2O | | | x | Balsamo *et al* (2011) |
| JULES | E2O | | | x | Best *et al* (2011) |
| LISFLOOD | E2O | | | x | Burek *et al* (2013) |
| ORCHIDEE | E2O | | | x | Krinner *et al* (2005) |
| PCR-GLOBWB | E2O | | | x | Wada *et al* (2014) |
| SURFEX-TRIP | E2O | | | x | Oki and Sud (1998) |
| W3RA | E2O | | | x | van Dijk *et al* (2013) |
| WaterGAP3 | E2O | | | x | Eisner (2016) |
| **CLM-CROP**[a] | ISIMIP agriculture | | | x | Drewniak *et al* (2013) |
| **EPIC-BOKU**[a] | ISIMIP agriculture | x | x | x | Williams (1995), Izaurralde *et al* (2006) |
| **EPIC-IIASA**[a] | ISIMIP agriculture | | | x | Balkovič *et al* (2014) |
| **EPIC-TAMU**[a] | ISIMIP agriculture | | | x | Kiniry *et al* (1995) |
| **GEPIC**[a] | ISIMIP agriculture | | | x | Liu *et al* (2007), Folberth *et al* (2012) |
| **LPJ-GUESS**[a] | ISIMIP agriculture | x | x | x | Lindeskog *et al* (2013) |
| **LPJmL**[a] | ISIMIP agriculture | | x | x | Bondeau *et al* (2007) |
| **ORCHIDEE-CROP**[a] | ISIMIP agriculture | x | | x | Wu *et al* (2016) |
| **PAPSIM**[a] | ISIMIP agriculture | | | x | Elliott *et al* (2014) |
| **PDSSAT**[a] | ISIMIP agriculture | | | x | Elliott *et al* (2014) |
| **PEGASUS**[a] | ISIMIP agriculture | | | x | Deryng *et al* (2014) |
| **PEPIC**[a] | ISIMIP agriculture | x | | | Liu *et al* (2016) |
| **CARAIB**[c] | ISIMIP biomes | x | x | x | Dury *et al* (2011) |
| **DLEM**[c] | ISIMIP biomes | x | x | x | Tian *et al* (2015) |
| **JULES-B1** | ISIMIP biomes | x | x | | Clark *et al* (2011) |
| **LPJ-GUESS** | ISIMIP biomes | x | x | x | Smith *et al* (2014) |
| **ORCHIDEE** | ISIMIP biomes | x | x | x | Krinner *et al* (2005) |
| **VEGAS**[c] | ISIMIP biomes | x | | x | Zeng *et al* (2005) |
| **VISIT** | ISIMIP biomes | x | x | x | Ito and Inatomi (2012) |
| **CLM** | ISIMIP global water | x | x | x | Leng *et al* (2015) |
| **DBH** | ISIMIP global water | x | x | x | Tang *et al* (2007) |
| **H08** | ISIMIP global water | x | x | x | Hanasaki *et al* (2008) |
| **JULES-W1** | ISIMIP global water | x | x | x | Best *et al* (2011) |
| **LPJmL** | ISIMIP global water | x | x | x | Sitch *et al* (2003) |
| **MATSIRO** | ISIMIP global water | x | x | x | Pokhrel *et al* (2015) |
| **MPI-HM** | ISIMIP global water | x | x | x | Stacke and Hagemann (2012) |
| **PCR-GLOBWB** | ISIMIP global water | x | x | x | Wada *et al* (2014) |
| **SWBM** | ISIMIP global water | x | x | x | Orth and Seneviratne (2015) |
| **VIC** | ISIMIP global water | x | x | x | Liang *et al* (1994) |
| **WaterGAP2** | ISIMIP global water | x | x | x | Müller Schmied *et al* (2016) |
| ERA-Interim/Land | Land Reanalyses | | | | Balsamo *et al* (2015) |
| MERRA-2 | Land Reanalyses | | | | Reichle *et al* (2017) |
| LandFlux-EVAL[b] | Composite[b] | | | | Mueller *et al* (2013) |

[a] ET estimates from crop model simulations participating in the ISIMIP2a agriculture sector are performed as pure crop runs only reporting growing season ET without the fallow period for a specific crop type and irrigation scenario (see more details in section 2.1).

[b] We distinguish here between four data subsets: (1) diagnostic datasets only (denoted as LFE DIAGNOSTIC), (2) land surface model (LSM) simulations (LFE LSM), (3) reanalysis datasets (LFE REANALYSES), and (4) all datasets combined (LFE ALL).

[c] As naturalized runs are not available for these biomes models, we have used varsoc simulations instead (i.e. simulations for which climate, population, gross domestic product, land use, technological progress and other parameters vary over the historical period).

in table 1, ISIMIP2a crop model simulations are provided as pure crop runs (i.e. assuming a specific crop is growing all over the world), implying that the ET output from these models corresponds to the amount of water evaporated or transpired from a specific crop type under a given irrigation scenario. Among all crop model simulations available within the ISIMIP2a framework, we focus here on ET from non-irrigated maize crops (denoted as $ET_{maize}$), as this crop type is among the most dominant crop types in the regions of interest (see section 3.2 and figure A1) and its growing season matches the summertime growing season on the Northern Hemisphere (see figure A2), making the comparison of $ET_{maize}$ with $ET_{tot}$ more straight-

forward. Note that additional preprocessing steps are required prior to performing these comparisons (listed in appendix B).

To additionally include simulations used in the scope of other model inter-comparison projects than ISIMIP, we also include estimates from simulations used in the E2O Water Resources Reanalysis v.1 (Schellekens *et al* 2017). Please refer to the latter publication for a short description of each of the contributing (land surface, global hydrological and simple water balance) models. Note that SWBM is also part of E2O but has been excluded from this ensemble, as its version is identical to the one of SWBM provided in ISIMIP2a. PCR-GLOBWB and Water GAP are also part of both inter-comparison projects, but kept in both ensembles due to differences in the model versions.

Besides the aforementioned simulations, we also analyse ET output from two major land reanalysis products: ERA-Interim/Land and MERRA-2. ERA-Interim/Land has been selected, as it is commonly used as a reference for quantifying land surface conditions (89 citations as of September 14, 2017, according to Web of Science). MERRA-2 is the most recent reanalysis advancement from NASA that uses refined precipitation corrections Reichle *et al* (2016). Although ET from MERRA-2 is known to have an anomalously high share of bare soil evaporation Schwingshackl *et al* (2017), the global average of total ET is well within the range of ET from other reanalyses Bosilovich *et al* (2016), making it a sufficiently good candidate for our analysis.

### 2.2. Diagnostic estimates

We use an ensemble of recent diagnostic datasets suitable for identifying potential biases in model-based estimates of ET (see table 1). This ensemble consists of some well established ET estimates from the recent past (a subset of those diagnostic datasets has also been used to generate the ensemble of diagnostic ET in LFE, see section 2.3), but also includes more recent datasets such as GLEAM v3.1 Martens *et al* (2016). Please refer to the references listed in table 1 for further details on the individual datasets. Please note that the version of Fluxnet-MTE employed here uses a modified temperature and precipitation forcing compared to the previous version presented in Jung *et al* (2009).

### 2.3. Composite estimates: LandFlux-EVAL

LandFlux-EVAL (LFE) is an ensemble based land ET product that itself is based on individual diagnostic, model and reanalysis products available in the early 2010s Mueller *et al* (2013), which we use here as a reference dataset (96 citations as of September 14, 2017, according to Web of Science). Although we cannot argue that LFE is free of biases, we can make the conservative assumption that its ensemble statistics (in particular the provided quantile statistics) are a suitable estimate for quantifying the probable range of ET over land. Please note that some of the diagnostic and model datasets used in the LFE ensemble are also included as individual datasets in this analysis, and hence there is some degree of dependency between the diagnostic (model based) estimates and both LFE DIAGNOSTIC (LFE LSM) and LFE ALL.

**Table 2.** List of prominent drought events that have occurred during the study period.

| Region | Drought event year | Citations |
|---|---|---|
| | 1987 | Trenberth *et al* (1988) |
| Great Plains | 2002 | Pielke *et al* (2005) |
| | 2012 | Hoerling *et al* (2013) |
| Central Europe | 2003 | Schär *et al* (2004) |
| | | García-Herrera *et al* (2010) |
| Western Russia | 2010 | Barriopedro *et al* (2011) |
| | | Hauser *et al* (2016) |

## 3. Methods

### 3.1. Remapping

In order to allow an inter-comparison of all datasets (mostly available at 0.5° resolution), we have bilinearly interpolated all input data to a 1° × 1° regular latitude-longitude grid (which corresponds to the spatial resolution of LFE). As LFE can be interpreted as an ET reference product (see section 2.3), we favoured to proceed with this lower resolution.

### 3.2. Study areas

Besides global land (excluding Antarctica), we focus on the following regions (see table 2): (1) the Great Plains region (spatial delineation according to the USGS physiographic divisions of the conterminous US, https://water.usgs.gov/lookup/getspatial?physio), (2) Central Europe (37°N–53°N and 1°W–20°E, representing the European land area with the highest JJA temperature anomalies in summer 2003 according to Schär *et al* (2004)) and (3) western Russia (50°N–60°N and 35°E–55°E, according to Dole *et al* (2011)). These regions have been selected, as they contain large areas of agricultural land that has been affected by severe heat waves and droughts during the time period covered by most of the assessed datasets. The spatial extent of the study areas is also indicated in figures A2 and A1. For each of the regions, area averages of ET were derived by weighting each grid cell by its land area (based on ISLSCP II Global Population of the World, Balk *et al* 2010).

### 3.3. Cluster analysis

We apply two distinct approaches to determine the overall level of similarity between the individual datasets focusing on (a) a combination of spatial and temporal variability (replicating the method described and applied in Sanderson and Knutti (2012) and Sanderson *et al* (2015) for the univariate case) and (b) spatial variability alone (following Mueller *et al* 2011). For both approaches, we first create a subset only containing data from the period of temporal

overlap (1989–2005) using all datasets except from GETA 2.0, MODIS Global ET and WECANN (which are not of sufficient length). If any of the grid cells (from any dataset and time step) contain missing data, values corresponding to those grid cells are removed from all datasets for all time steps. From the $m = 75$ ($m = 89$ for $ET_{maize}$) datasets remaining, we then proceed as follows:

In approach (a), we first reform the elements of the four-dimensional input field (latitude, longitude, time and dataset) to a two-dimensional matrix X of size $m$ by $n$ (where $n$ corresponds to the number of all non-missing observations) and transform this into anomalies (based on all time steps remaining) $\Delta X$. To only preserve the dominant modes of ensemble variability, a singular value decomposition (SVD) is performed on $\Delta X$ and truncated to $t = 9$ modes (which is well within the range of suitable truncation values found in Sanderson *et al* (2015)). The corresponding loadings matrix U (size $m$ by $t$) can now be used to derive the difference matrix $D_{svd}$ by calculating pairwise Euclidean distances as

$$\delta_{ij} = \left\{ \sum_{l=1}^{t} [\mathbf{U}(i,l) - \mathbf{U}(j,l)]^2 \right\}^{1/2}.$$

In approach (b), we simply calculate the temporal average (denoted 'tavg') and reform the result to a two-dimensional matrix Y of size $m$ by $p$ (where $p$ corresponds to the number of remaining grid cells, which is the same for all datasets). We then calculate pairwise Euclidean distances $\delta_{ij}$ in between datasets $i$ and $j$, resulting in the difference matrix $D_{tavg}$ (dimensioned $m$ by $m$):

$$\delta_{ij} = \left\{ \sum_{l=1}^{p} [\mathbf{Y}(i,l) - \mathbf{Y}(j,l)]^2 \right\}^{1/2}.$$

The distance matrices $D_{svd}$ and $D_{tavg}$ are then subjected to hierarchical clustering by applying the *hclust* function in R statistics software with Ward's clustering criterion (Ward 1963, ward.D2). The results are visualized in R using the *Complex Heatmap* package (version 1.14.0, available at https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html). Clusters arevisually highlighted using a fixed threshold corresponding to the 99th percentile of the distances to enhance comparability of the results. To compare these clusters against the influence of the assessed factors (i.e.the data category, forcing, number of soil layers, ET scheme and model choice) on overall ET variability, we also assess the fraction of variation explained by each of those factors. This is achieved by means of applying a permutational multivariate analysis of variance (PERMANOVA, Anderson 2001) on the Euclidean differences $D_{svd}$ and $D_{tavg}$.

## 4. Results and discussion

### 4.1. Spatio-temporal patterns of global land ET

Figure 1 shows global patterns of time averaged total ET (i.e. excluding crop models) grouped by data category. Spatial patterns of the ensemble averages are mostly similar among data categories and delineate the average location of the governing hydroclimatic regimes, such as the Intertropical Convergence Zone. The ensemble spread (as expressed by the IQR, i.e. the difference between the 75th and 25th percentile) in tropical rainforest regions is generally largest for the ISIMIP simulations and LFE ALL, whereas the uncertainties (in absolute values) in this region are considerably less pronounced in the diagnostic and E2O ensemble. Similarly, the spatial patterns of the relative ensemble spread (shown by means of the quartile coefficient of dispersion, QCD, which is a relative measure of dispersion, Bonett 2006) in the diagnostic ensemble are very well represented by both ISIMIP sectors and by LFE ALL. Although the magnitude of the uncertainties in the diagnostic ensemble is smaller than in the ISIMIP sectors, it is still higher than the relative ensemble spread in E2O. Note, however, that these small inter-model differences are linked to the fact that the E2O models are all forced by WFDEI, while the ISIMIP ensembles consist of three different forcings.

The results of the SVD-based cluster analyses are shown in figures 2 ($ET_{tot}$) and figure 3 ($ET_{maize}$). As the associated clusters explain most of the variability among the assessed datasets (i.e. incorporating variability in both their spatialand temporal domains), we treat those as our main results, but occasionally draw comparisons to the clustering results based on time-averaged ET (visualized in figures A3 and figure A4). In the following, we highlight important details of the cluster diagrams by discussing individual parameters potentially having a measurable influence on the differences among the analysed ensemble.

At first glance, we notice a few apparent outliers in the displayed difference matrices, most strikingly for WB-MTE ($ET_{tot}$), PM-MU CSIRO, MPI-HM and the crop models LPJ-GUESS and ORCHIDEE-CROP ($ET_{maize}$ only). As the differences among the contributing models are affected not only by the modelling structure but also the calibration process taken, it is not surprising to see such apparent differences between the non-calibrated models LPJ-GUESS and ORCHIDEE-CROP and all other datasets (the less pronounced signal for ORCHIDEE-CROP is potentially due to the fact that this model includes Nitrogen cycling). We also see that the long-term average global land mean *ET* of these models is not an outlier, suggesting that the differences are mainly due to anomalous patterns in the spatial domain. This is also supported by the clustering results of time-averaged ET, where outliers in the dendrogram mostly coincide with anomalies in the long-term mean (e.g. WANG-ET for $ET_{tot}$, figure A3 or PEGASUS for $ET_{maize}$, figure A4).

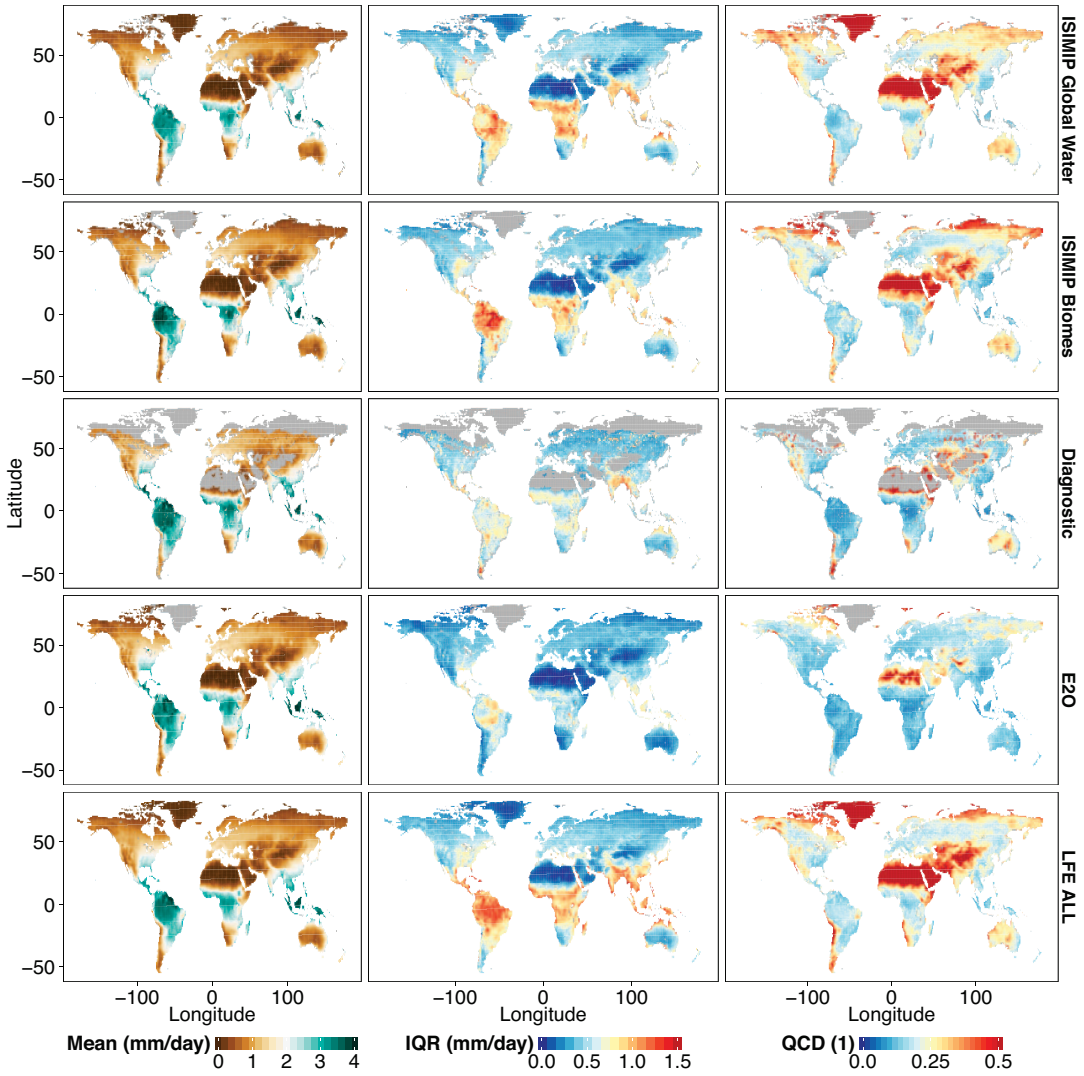


**Figure 1.** Temporal averages (1989–2005 time period) of monthly ensemble means (left), ensemble IQRs (middle) and ensemble QCDs (right) of $ET_{tot}$ grouped by data category (rows). The diagnostic ensemble is based on all diagnostic datasets except GETA 2.0, MODIS Global ET and WECANN (which have been excluded due to insufficient length). ERA-Interim/Land and MERRA-2 are not shown.

The meteorological forcing has, in general, a very noticeable impact on *ET* estimates when considering variability across both space and time. There are a number of clusters whose members are almost exclusively based on the same forcing (e.g. all PGMFD-forced simulations form a single cluster for $ET_{tot}$; cluster 2 in figure 2), underlining the dominant role of this parameter. However, the forcing apparently only has very minor influence on the spatial variability of time-averaged $ET_{tot}$ (figure A3), indicating that most of the differences in between the employed forcing datasets are due to differences in the temporal evolution of ET. For $ET_{maize}$, there are a few more models for which the forcing dominates the differences, arguably due to a higher sensitivity of these models to diurnal forcings and other parameters which play a more important role when considering crop-specific ET aggregated over growing seasons.

Similarities among the individual datasets can also be reasonably well explained by their data category. While ET simulations from the ISIMIP global water and biomes models are very similar, crop models show substantial differences to all other realizations (the majority of the crop model simulations are members of the same cluster; cluster 3 in figure 3), reflecting the missing (or only rudimentary) representation of crops in the water and biomes models. E2O simulations also share most of their space-time variability (all but one of the E2O simulations fall into the same cluster; cluster 4 in figure 2 and cluster 5 in figure 3). This could be due to stronger similarities among the models participating in this project. Diagnostic datasets mostly show similarities in the spatial variability of the time averages (for $ET_{tot}$ in figure A3, all but two diagnostic datasets are associated with the same cluster). It is also noticeable that ET diagnostics from the LandFlux project (i.e. PM-MU LANDFLUX, PT-FI LANDFLUX and SEBS LANDFLUX) form their own cluster in the spatio-temporal dendrogram (most pronounced in cluster 7, figure 2). For $ET_{tot}$, we also note

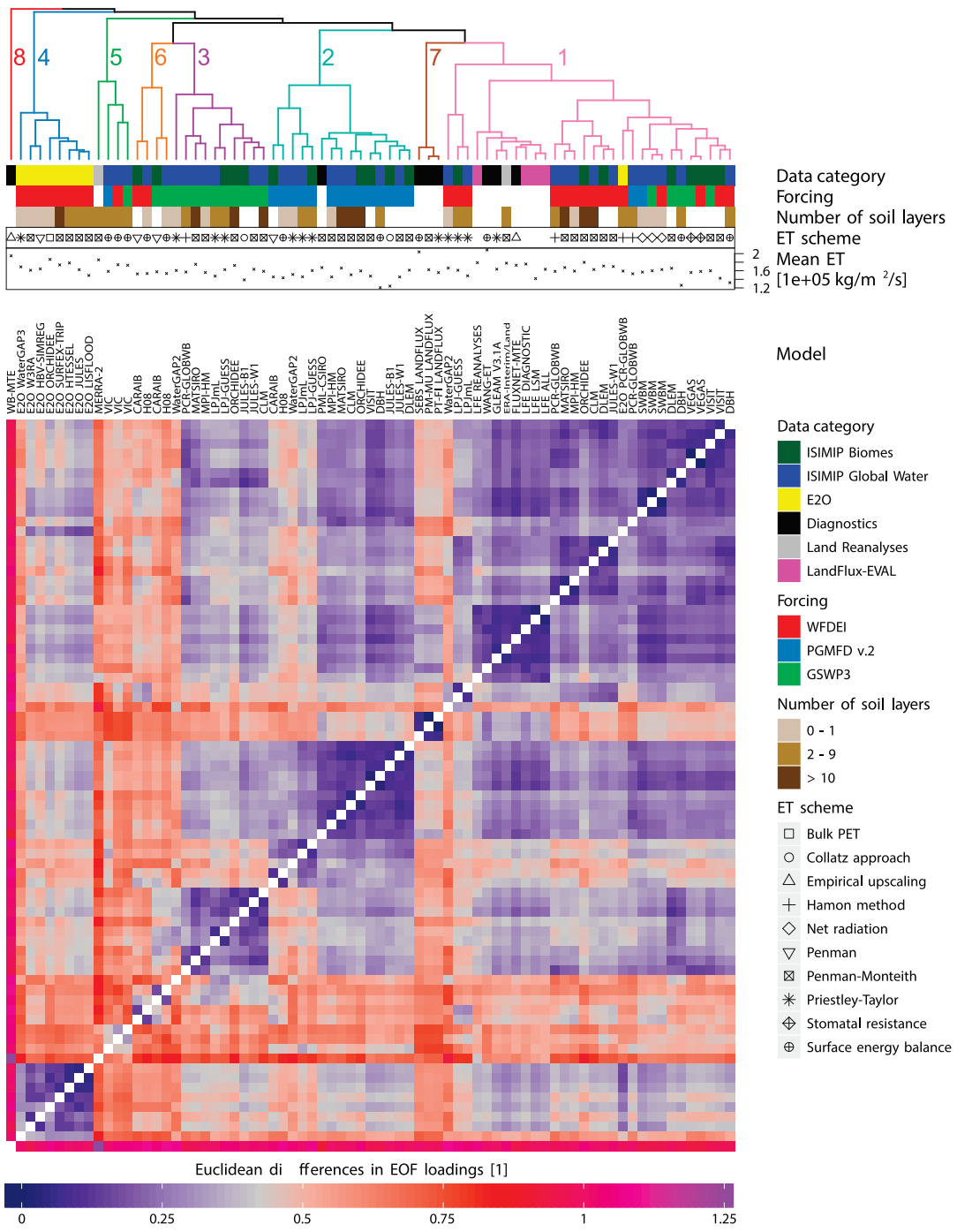


**Figure 2.** (Previous page) Euclidean differences based on $D_{svd}$ (capturing spatio-temporal variabilities) for global land $ET_{tot}$ (coloured matrix). Dataset names are indicated by the column names on top of the matrix (ISIMIP model simulations are printed in bold). For each dataset, the coloured or symbolized rows on top of the matrix indicate the associated data category, the meteorological forcing, the number of soil layers, the ET scheme and the time-average of global land mean ET of each dataset. White boxes or missing symbols indicate that an association is not possible. The dendrogram on the top displays the similarity among the datasets. Colours and numbers indicate clusters (ordered by their size) of datasets with particularly small differences among its members.

that the LFE products are quite similar to most of the diagnostic datasets (sub-tree of cluster 1 in figure 2), adding support that LFE can indeed be used as a reference.

The other two parameters displayed in the cluster diagrams are the (binned) number of soil layers (water and crop models only) and the ET parametrization scheme. These parameters apparently only play a minor role for the clustering. While there are still a couple of instances where the ET scheme seems to contribute to small differences among the datasets (e.g. sub-trees of clusters 1, 2 and 4 in figure 2 indicate that datasets using the Penman-Monteith formulation show particularly little differences), there is no apparent influence of the number of soil layers. It must also be noted that there are many cases where both the ET scheme and the model are the same (e.g. cluster 5 in figure 2), suggesting that (in these cases at least) the associated

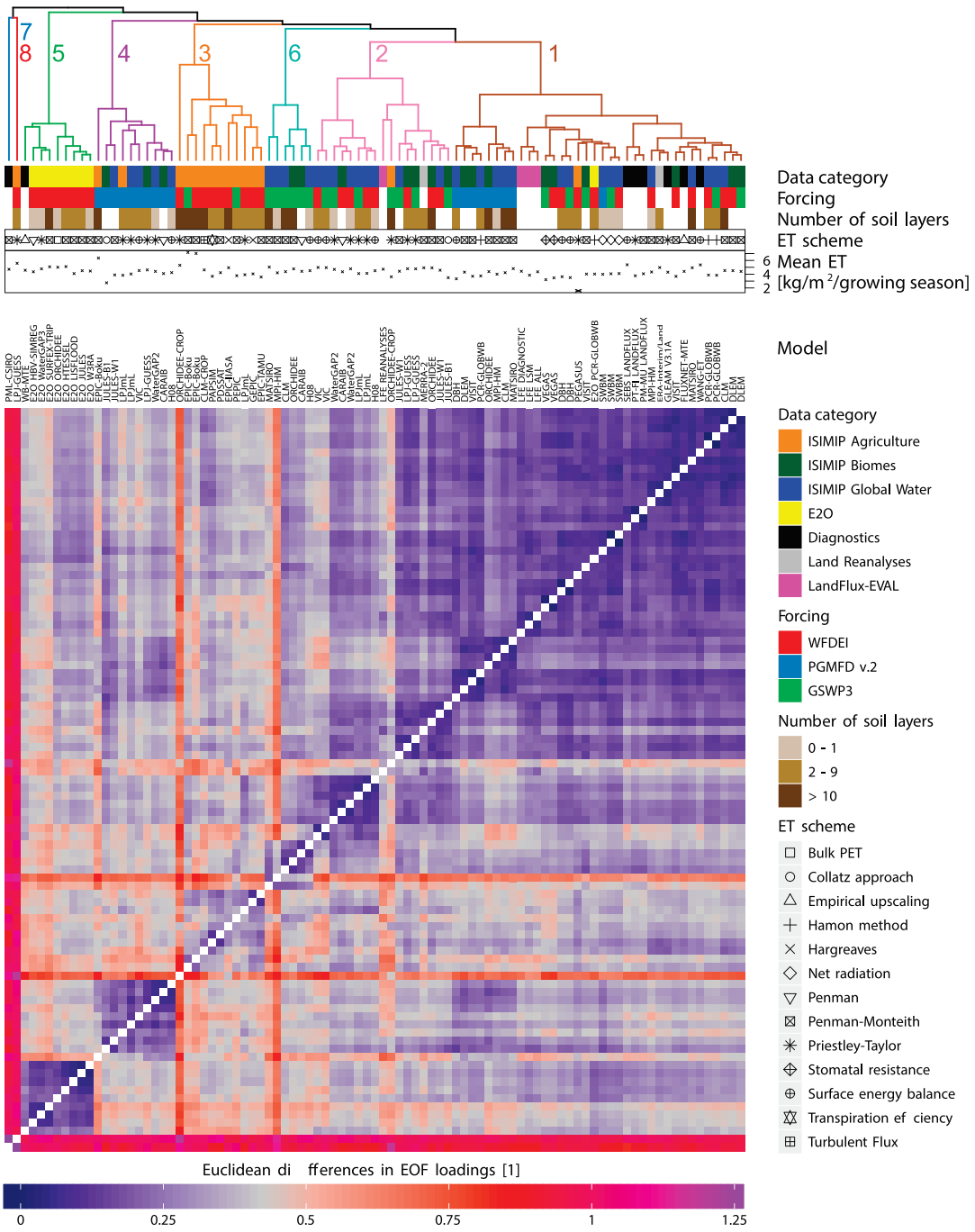


**Figure 3.** As in figure 2, but for global land $ET_{maize}$ (note the different units).

small differences are due to the identity of the model. These findings are in contrast to what we would have expected from the literature, which suggests that ET schemes may explain a substantial part of the overall variance in ET (see section 1), whereas our results rather suggest that the forcing dominates this variance.

Besides looking at the individual clusters, it is also important to assess the fractional variance in *ET* for each of the factors considered (figure 4). In all cases, the assessed factors are capable of explaining more than 90% of the variabilities in the distance matrices ($D_{svd}$ and $D_{tavg}$), and the identified proportions of explained variance are all statistically significant ($p < 0.01$). Irrespective of the distance metric and the type of ET estimate, differences in the forcing, model choice and ET scheme together account for more than half of the variabilities. In fact, the combined effect of model choice and ET schemes alone can explain at least 48% of the variabilities (which is again in line with previous findings that stress the importance of ET schemes). When considering $D_{svd}$, the forcing plays an important (for $ET_{tot}$ even the leading) role, whereas differences in $D_{tavg}$ can mainly be explained by the choice of the model. In spite of its minor role for the cluster analysis, the ET scheme can explain more than a quarter of the variabilities.

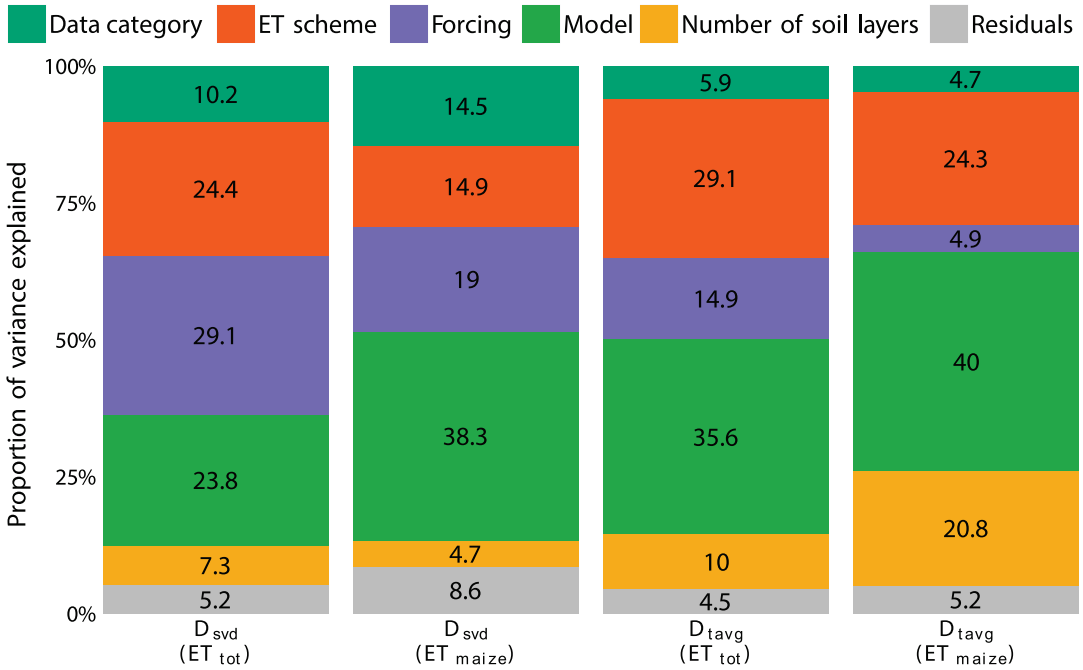


**Figure 4.** Fraction of variance in $D_{svd}$ (representing spatio-temporal variabilities) and $D_{tavg}$ (representing spatial variabilities) for both $ET_{tot}$ and $ET_{maize}$ explained by different factors (coloured shading and numbers [%]). Also shown is the proportion of unexplained variance (grey shading).


**Figure 5.** Time series of annual mean ET averaged over global land, the Great Plains, Central Europe and western Russia (rows). The first and second columns show ensemble medians and IQRs of the four largest clusters (numbered 1–4; names in brackets correspond to the dominant meteorological forcing or data category found within each cluster, as displayed in figure 2). Also shown is the median of LFE ALL. The rightmost column shows area averages of individual diagnostic (black, purple, blue and green) and reanalysis datasets (brown), and the associated IQR (grey shading). The long-term average ET estimate from GETA 2.0 is shown as a horizontal purple line. Drought events in the different regions are marked by vertical red lines.

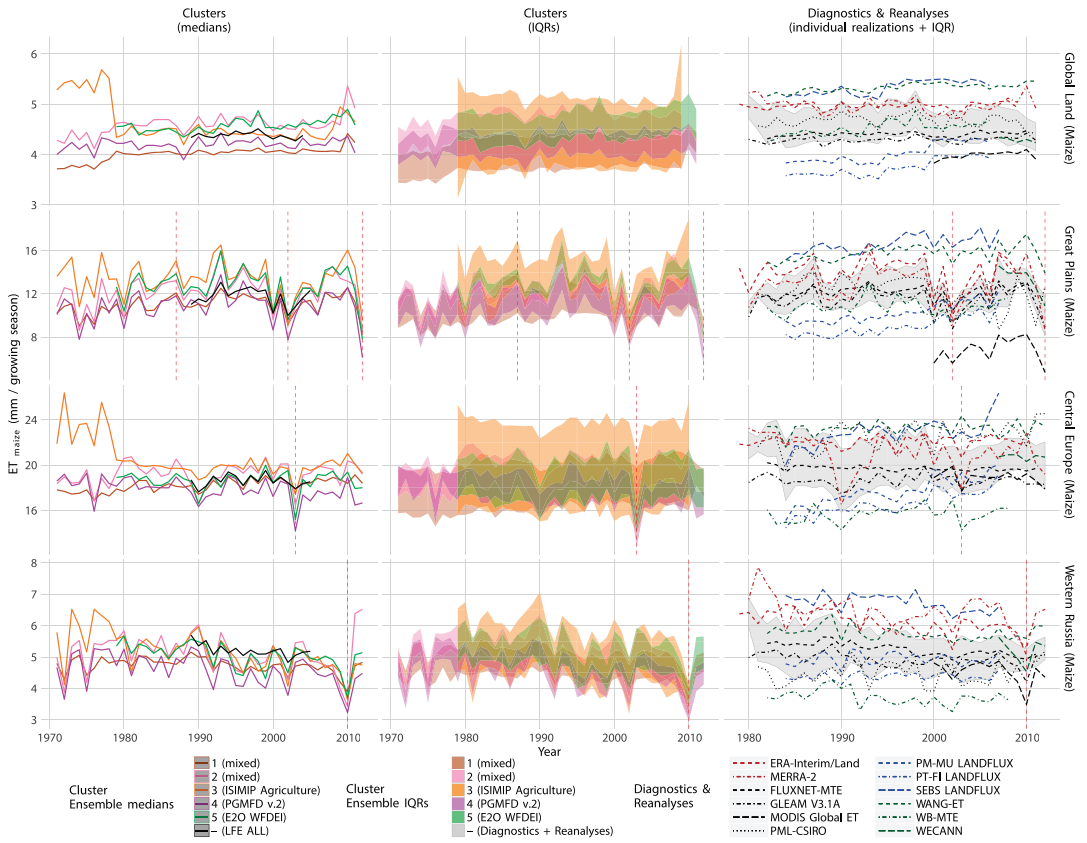


**Figure 6.** Like figure 5, but for fractional ET for non-irrigated maize crops, also including crop models from the ISIMIP2a agriculture sector. The last year of the global land estimates is excluded here, as it also contains ET from maize crops within the Southern Hemisphere, where the respective growing seasons usually end in the next calendar year (see figure A2).

We now have a good overview of the structural differences among the assessed ensemble of datasets. Based on this information (using the largest sub-ensembles shown in figures 2 and figure 3), we can now go ahead and investigate differences among the clusters with respect to both the temporal evolution of ET and the representation of extreme events.

### 4.2. Temporal evolution of global and regional averages of ET

Figures 5 and figure 6 present global and regional time series of both $ET_{tot}$ and $ET_{maize}$, respectively. Among the diagnostics- and reanalysis-based estimates we see a few noticeable outliers for both $ET_{tot}$ and $ET_{maize}$, most notably SEBS LANDFLUX (for which the global and regional averages exceed the IQR of the entire time span). The magnitude of the spread (IQR) among those estimates is comparable to the magnitude of the spread of the displayed cluster-based sub-ensembles. In the global domain, the diagnostics- and reanalysis-based estimates of $ET_{tot}$ are slightly positively biased with respect to the other sub-ensembles shown. Those differences are most pronounced for cluster 2 (which is also located well below the other cluster ensembles and is also offset with respect to the median of LFE ALL), suggesting an underestimation of global land $ET_{tot}$, which is potentially related to the dominance of PGMFD-forced model simulations in this cluster. However, the respective sub-ensemble for $ET_{maize}$ (cluster 4 in figure 6) is well embedded in the overall spread for all regions considered. We also observe that crop models from the ISIMIP agriculture sector tend to disagree more on $ET_{maize}$ than datasets from the other clusters, although their median is well in agreement with the medians of the other sub-ensembles (cluster 3 in figure 6; note that this ensemble consists of only one model prior to 1979). Except from this, the medians of the clusters agree reasonably well, both in terms of their absolute magnitude (which is close to the median of LFE ALL) and their temporal variability.

The prominent jump in $ET_{tot}$ from 1978–1979 apparent for cluster 1 within the Central European and global domains (figure 5) is due to the switch in the forcing data from WATCH to WFDEI at this time (when considering a sub-ensemble of only WFDEI-forced simulations, this discontinuity becomes even more pronounced, not shown). This artificial change signal is caused by differences in the reanalysis product used to create these datasets (ERA-40 for WFD, ERA-Interim for WFDEI), and is mainly due to the revision in the average aerosol loadings in North Africa and Europe affecting the short-wave incoming solar radiation in these regions (Weedon *et al* 2014).

A particularly important aspect to assess in greater detail is the response of the different datasets and sub-ensembles to drought events in the study areas (drought

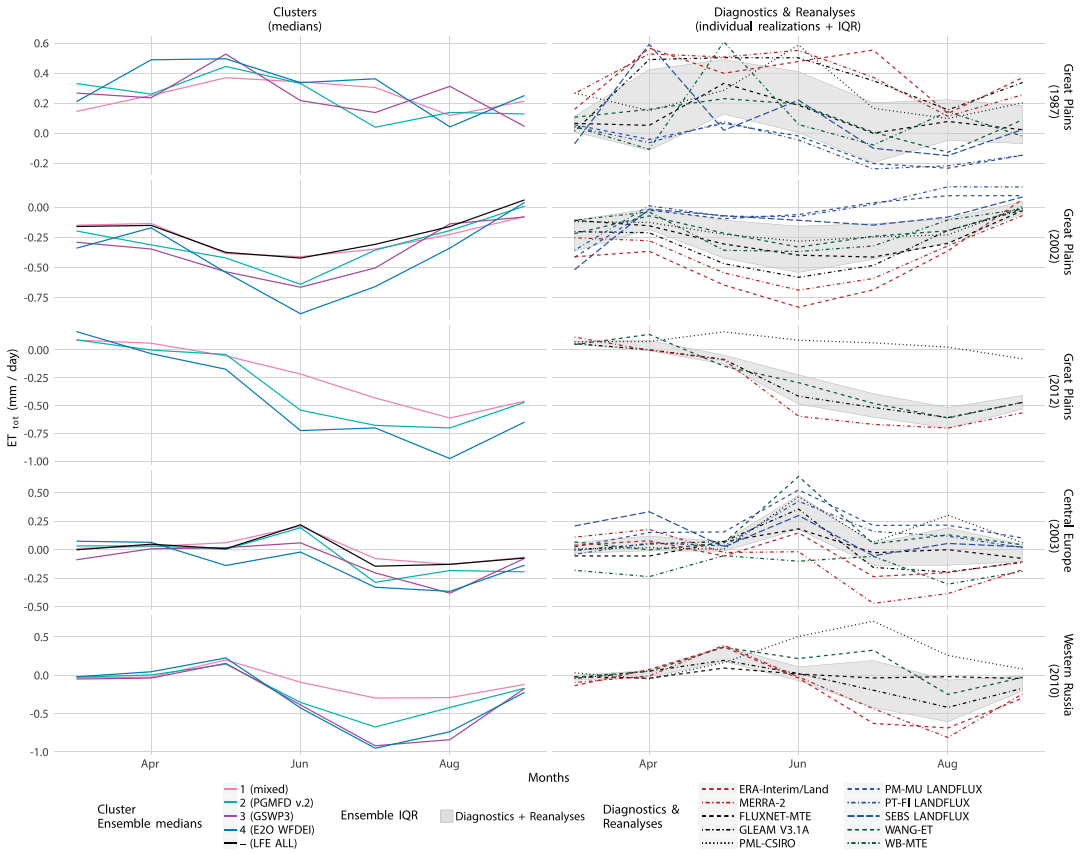


**Figure 7.** Like figure 5, but for monthly ET anomalies (reference period 1989–2005), excluding GETA 2.0, MODIS Global ET and WECANN. Cluster IQRs are not displayed for clarity. Individual time series are centered at the event years shown in the labels on the right-hand side of each sub-plot.

years indicated in rows 2–4 in figures 5 and figure 6; in addition, monthly time series for each event year are shown in figure 7). Irrespective of the region and event considered, month-to-month changes in ET anomalies in the displayed sub-ensembles are similar to the changes in the ensemble mean of the diagnostic estimates. However, we note that there is a tendency of the model-dominated sub-ensembles to exaggerate negative ET anomalies (relative to the diagnostic estimates), which could be related to overly strong soil water depletion in some of the model simulations, to models overestimating stomatal closure, to some models not representing irrigation, or to missing non-stomatal flux components.

For each of the considered regions, there has been at least one pronounced drought during the analysed time span (see table 2). While the 1987 Great Plains drought has no apparent influence on ET rates, the events in 2002 and 2012 are characterized by negative ET anomalies apparent in both the displayed clusters and in most of the diagnostic estimates (the anomaly signal is slightly more pronounced for $ET_{maize}$). In the case of western Russia, negative ET anomalies accompanying the 2010 heat wave are found in most of the diagnostic datasets, in the reanalyses, and most strikingly in the (model-dominated) clusters, which is in line with the ET signal discussed in Hauser *et al* (2016). For the European heatwave in 2003, the diagnostic estimates show a noticeable disagreement from the models and reanalyses. However, the month-to-month variations in ET anomalies from most of the datasets and simulations (figure 7) still resemble the event. The patterns agree both with theory Seneviratne (2012) and catchment-level observations Teuling *et al* (2013), as anomalously dry weather (accompanied by high net radiation and anomalously warm temperatures) commonly leads to an initial increase in ET, followed by a decrease later in the season (when soil moisture reaches critical levels). These temporal patterns can partly also be observed for the other events (although the initial positive ET anomalies are less pronounced or located earlier in the season).

## 5. Conclusions

In this paper we have shown that evapotranspiration simulated by models in the global water, biomes and agriculture sectors of ISIMIP2a is prone to substantial uncertainties. By means of applying cluster analyses on Euclidean differences in the spatial and spatio-temporal domains (using a large ensemble consisting

**Table 3.** Uncertainties (standard deviations over selected sub-ensembles) of global mean ET averaged over 1989–2005.

| Cluster number | $ET_{tot}$ (mm day$^{-1}$) | $ET_{maize}$ (mm/growing season) |
|---|---|---|
| – (all datasets) | 0.15 | 0.75 |
| 1 | 0.15 | 0.67 |
| 2 | 0.13 | 0.44 |
| 3 | 0.11 | 1.31 |
| 4 | 0.11 | 0.94 |
| 5 | — | 0.36 |

of both the ISIMIP simulations and various other datasets at monthly resolution), we have found remarkable similarities within specific sub-ensembles. This clustering could be attributed to several governing factors, including, in order of relative importance, (1) the meteorological forcing used to drive the model simulations (particularly relevant when considering variability in both space and time), (2) the data category associated with each dataset (e.g. we observe a pronounced clustering of E2O simulations), (3) the model choice, (4) the ET scheme (e.g. there is some clustering of models using the Penman-Monteith formulation) and, apparently less important, (5) the number of soil layers. However, the partitioning of the spatial and spatio-temporal variances in the Euclidean distance matrices ($D_{tavg}$ and $D_{svd}$) reveals that the ET scheme (and partly also the number of soil layers) still explains a relatively large fraction (together 20% to 39%) of the variance. Although the model choice explains a major fraction of the spatial and spatio-temporal variance, the forcing plays an even more important role for the clustering of the SVD-based (i.e. spatio-temporal) differences in $ET_{tot}$. This underlines that besides further model improvements it is at least equally important to further reduce uncertainties in the forcing datasets.

Mean uncertainties among the assessed sub-ensembles are lowest for GSWP3-forced ISIMIP simulations ($ET_{tot}$) and for EartH2Observe simulations (both $ET_{tot}$ and $ET_{maize}$; see table 3). However, this does not imply unbiasedness of the corresponding ensemble averages. Irrespective of the substantial uncertainties, the global and regional averages of the cluster-based sub-ensembles are in reasonable agreement to each other. This also holds for crop-specific ET, where the investigated crop models have a similar median tendency but larger inter-model spread. Considerable differences to the reference ET from LandFlux-EVAL and also to the other sub-ensembles have only been found for the subset of ISIMIP models driven by PGMFD v.2 (cluster number 2, mainly emerging in the global land average).

We have further assessed the representation of selected droughts and heat waves in the identified sub-ensembles. We could demonstrate that most of the assessed datasets show the anticipated signal (i.e. negative ET anomalies that emerge after an initial surplus in ET), although the magnitude of these anomalies is usually of the same order as the magnitude of the spread among the estimates.

It is tempting to use our results to assign weights to individual datasets using e.g. their bias with respect to parameters describing the central tendency of the entire ensemble. However, as we cannot ensure the unbiasedness of the ensemble mean, we suggest not to do this. As a recommendation, it seems suitable to rather choose the whole ensemble or a representative sub-ensemble.

Note that this work should be regarded as a first step towards more detailed analyses. Further research is needed in particular to address the influence of different crop types (e.g. wheat, rice or soy beans) on simulated ET. Another challenge is to assess the interplay of land use types and ET rates or extremes in the hydrological cycle. However, such questions can currently not be well addressed in a comprehensive, cross-sectoral way due to the substantial differences in the nature of the contributing models, and hence further advances towards reducing the complexity of inter-sectoral comparisons are needed.

## Acknowledgments

This research was funded by the European Research Council DROUGHT-HEAT project (contract 617518). PEGASUS simulations were carried out by D. Deryng on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia. GL was supported by the Office of Science of the US Department of Energy as part of the Integrated Assessment Research Program. CLM water sector simulations were performed using PNNL Institutional Computing at Pacific Northwest National Laboratory. PNNL is operated by Battelle Memorial Institute for the US DOE under contract DE-AC05-76RLO1830. GPW was supported by the Joint DECC and Defra Integrated Climate Program - DECC/Defra (GA01101). J L was supported by the National Natural Science Foundation of China (41625001, 41571022) and partly supported by the Southern University of Science and Technology (Grant no. G01296001).

# Appendix A. List of figures

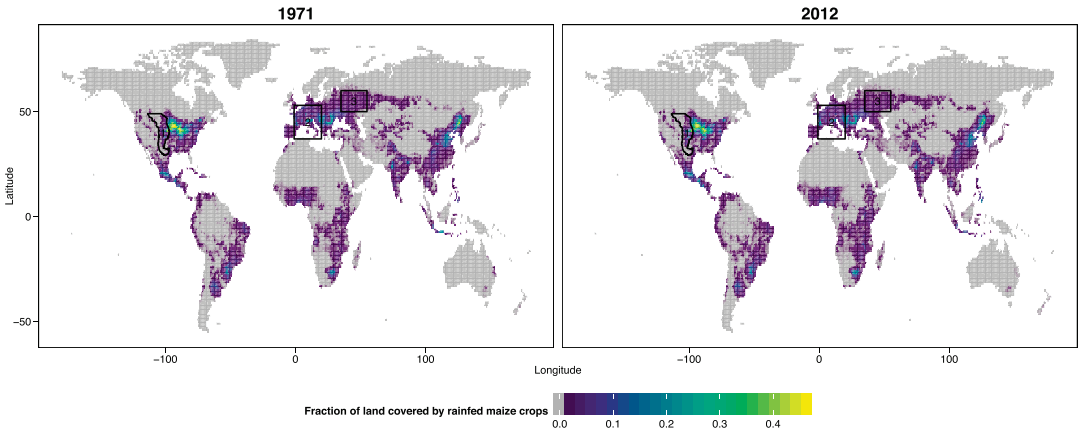


**Figure A1.** Fraction of land used for non-irrigated maize crops in 1971 and 2012, based on HYDE Hoekstra (1998) and appendix of Fader *et al* (2010) and MIRCA Portmann (*et al* 2010). The map also shows the regions of interest discussed insection section 4.2: 1 = Great Plains, 2 = Central Europe, 3 = western Russia.

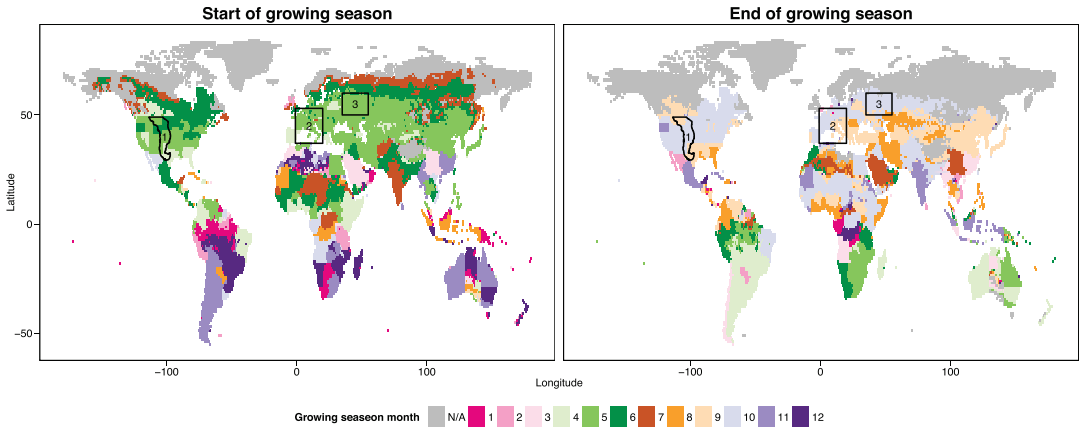


**Figure A2.** Growing season start end end months (rounded) for non-irrigated maize crops, based on (Sacks *et al* 2010, Portmann *et al* 2010 and Elliott *et al* 2015). The map also shows the regions of interest discussed in section section 4.2: 1 = Great Plains, 2 = Central Europe, 3 = western Russia.

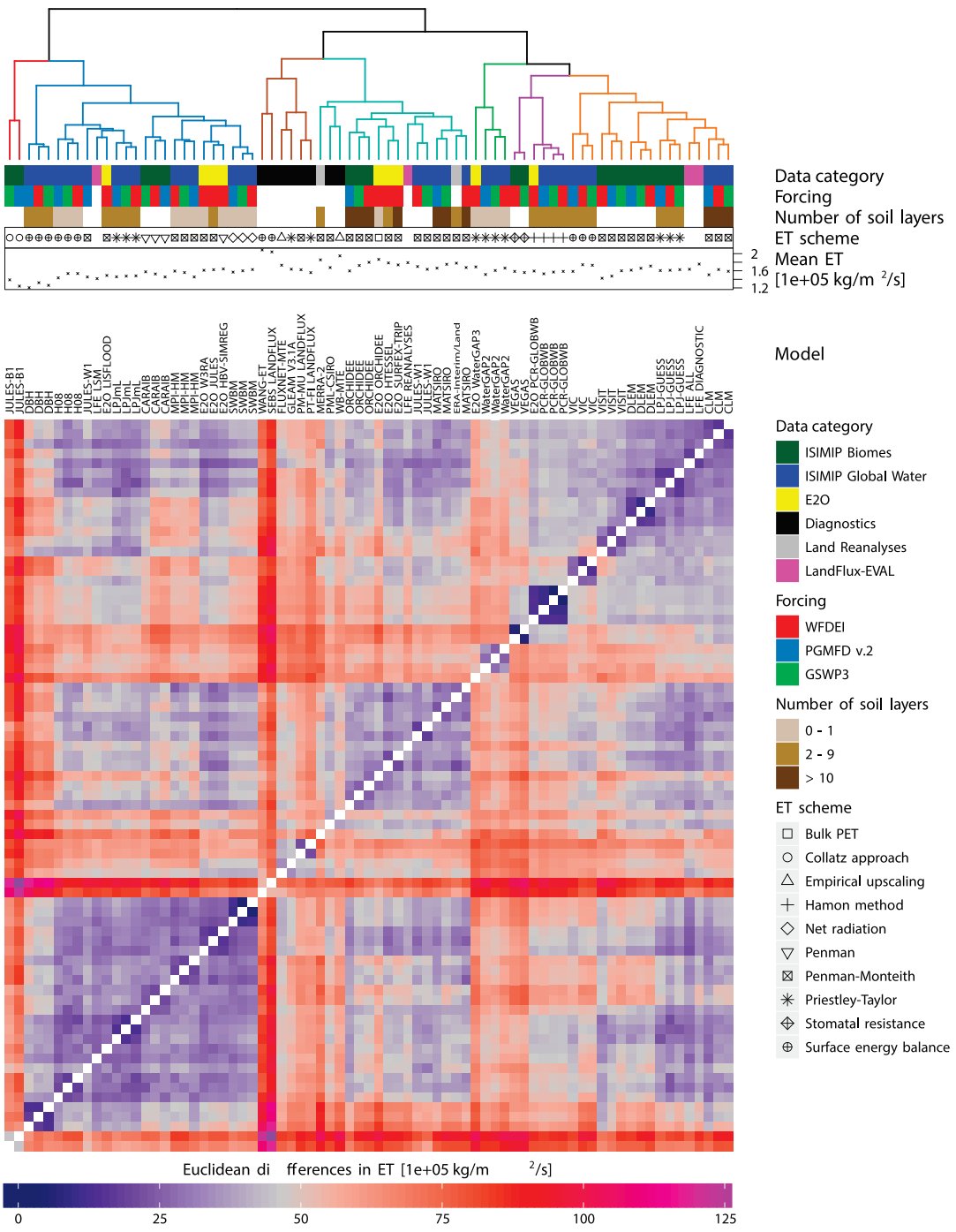


**Figure A3.** As in figure 2, but based on time-averaged $ET_{tot}$ ($D_{tavg}$).

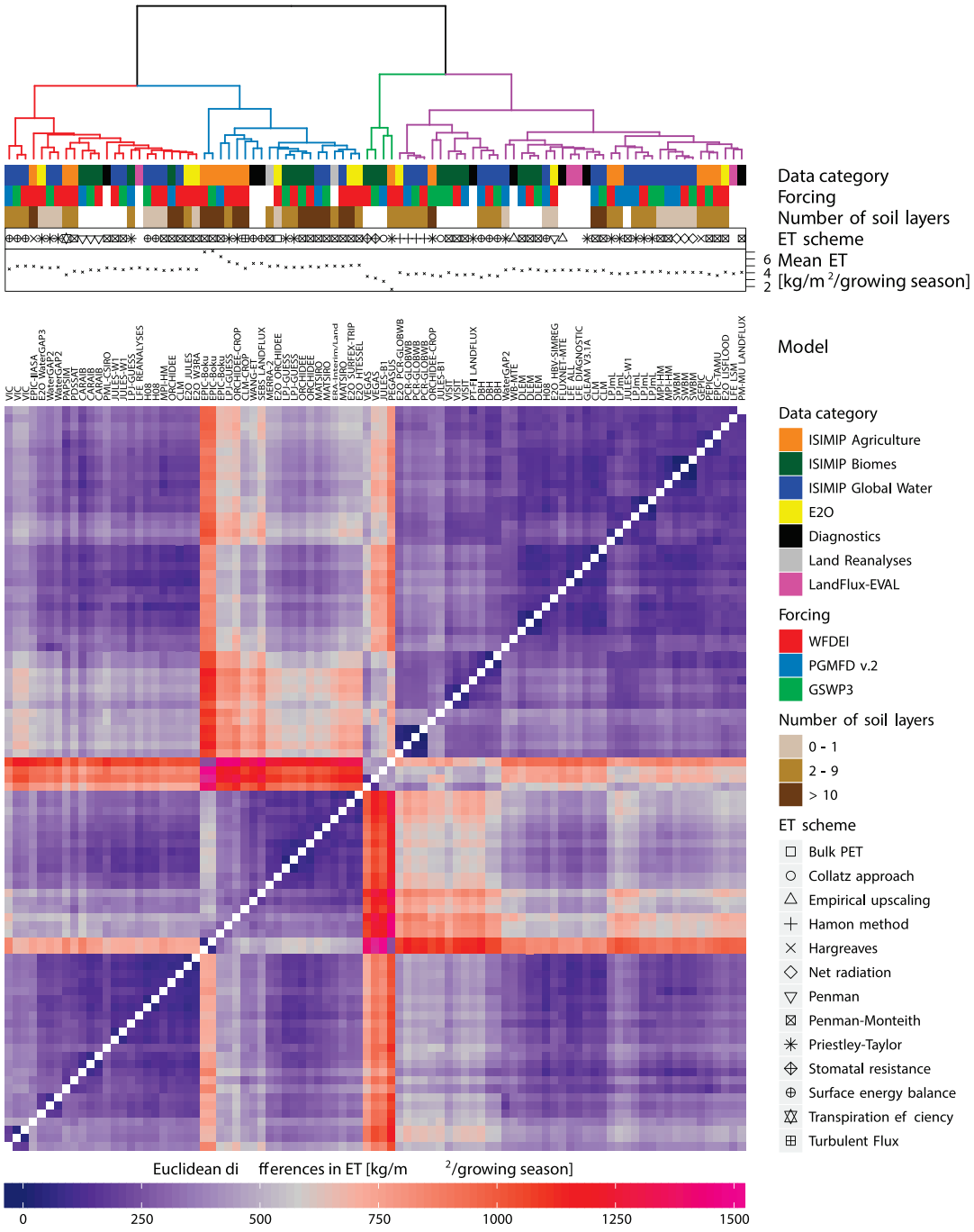


**Figure A4.** As in figure 3, but based on time-averaged $ET_{maize}$ ($D_{tavg}$).

## Appendix B. Data preprocessing for comparison with crop model output

In order to perform a cross-sectoral inter-comparison between $ET_{maize}$ and $ET_{tot}$, it was necessary to pre-process the original datasets. We first aggregated the monthly estimates (i.e. $ET_{tot}$) to match the growing season of the non-irrigated maize crops after rounding the respective growing seasons to full months (see figure A2). In a second step, all growing season-specific estimates of ET (i.e. including ET from the crop models) were multiplied by a gridded dataset of time-varying historical cropland patterns which is based on trends of agricultural land from HYDE Hoekstra (1998) and appendix of Fader *et al* (2010)) and present-day (year 2000) crop and irrigated areas from MIRCA (Portmann *et al* 2010, see figure A1). By these means, we ensured that an inter-comparisons of $ET_{maize}$ across all datasets is only performed for grid cells where non-irrigated maize crops are actually growing.

## ORCID iDs

Richard Wartenburger https://orcid.org/0000-0003-4470-5080
Sonia I Seneviratne https://orcid.org/0000-0001-5973-6862
Lukas Gudmundsson https://orcid.org/0000-0003-3539-8621
Xingcai Liu https://orcid.org/0000-0001-5726-7353
Christoph Müller https://orcid.org/0000-0002-9491-3550
Hannes Müller Schmied https://orcid.org/0000-0001-5330-9923
Thomas A M Pugh https://orcid.org/0000-0002-6242-7371
Qiuhong Tang https://orcid.org/0000-0002-0886-6699
Wim Thiery https://orcid.org/0000-0002-5183-6145
Yoshihide Wada https://orcid.org/0000-0003-4770-2539
Graham P Weedon https://orcid.org/0000-0003-1262-9984
Tian Zhou https://orcid.org/0000-0003-1582-4005

## References

Alemohammad S H, Fang B, Konings A G, Green J K, Kolassa J, Prigent C, Aires F, Miralles D and Gentine P 2016 Water, energy, and carbon with artificial neural networks (WECANN): a statistically-based estimate of global surface turbulent fluxes using solar-induced fluorescence *Biogeosci. Discuss.* **2016** 1–36

Ambrose S M and Sterling S M 2014 Global patterns of annual actual evapotranspiration with land-cover type: knowledge gained from a new observation-based database *Hydrol. Earth Syst. Sci. Discuss.* **2014** 12103–35

Anderson M J 2001 A new method for non-parametric multivariate analysis of variance *Austral Ecol.* **26** 32–46

Balk D, Deichmann U, Yetman F, Hall G, Collatz B, Meeson S, Los E B D C and Landis D 2010 Islscp ii global population of the world *Technical Report* (Oak Ridge, TN: ORNL DAAC)

Balkovič J, van der Velde M, Skalský R, Xiong W, Folberth C, Khabarov N, Smirnov A, Mueller N D and Obersteiner M 2014 Global wheat production potentials and management flexibility under the representative concentration pathways *Glob. Planet. Change* **122** 107–21

Balsamo G *et al* 2015 ERA-Interim/Land: a global land surface reanalysis data set *Hydrol. Earth Syst. Sci.* **19** 389–407

Balsamo G, Pappenberger F, Dutra E, Viterbo P and van den Hurk B 2011 A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle *Hydrol. Process.* **25** 1046–54

Barriopedro D, Fischer E M, Luterbacher J, Trigo R M and García-Herrera R 2011 The hot summer of 2010: redrawing the temperature record map of Europe *Science* **332** 220–4

Beck H E, van Dijk A I J M, de Roo A, Miralles D G, McVicar T R, Schellekens J and Bruijnzeel L A 2016 Global-scale regionalization of hydrologic model parameters *Water Resour. Res.* **52** 3599–622

Best M J *et al* 2011 The joint UK land environment simulator (JULES), model description—Part 1: Energy and water fluxes *Geosci. Model Dev.* **4** 677–99

Betts A K, Ball J H, Beljaars A C M, Miller M J and Viterbo P A 1996 The land surface-atmosphere interaction: a review based on observational and global modeling perspectives *J Geophys. Res. Atmos.* **101** 7209–25

Bondeau A *et al* 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance *Glob. Change Biol.* **13** 679–706

Bonett D G 2006 Confidence interval for a coefficient of quartile variation *Comput. Stat. Data An.* **50** 2953–7

Bormann H 2011 Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change atGerman climate stations *Clim. Change* **104** 729–53

Bosilovich M G, Robertson F R, Takacs L, Molod A and Mocko D 2016 Atmospheric water balance and variability in the MERRA-2 reanalysis *J. Clim.* **30** 1177–96

Burek P, Knijff van der J and Roo de A 2013 LISFLOOD, distributed water balance and flood simulation model revised user manual 2013 *Report* (Luxembourg: Publications Office of the European Union)

Chen T H *et al* 1997 Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes *J. Clim.* **10** 1194–215

Clark D B *et al* 2011 The joint UK land environment simulator (JULES), model description—Part 2: carbon fluxes and vegetation dynamics *Geosci. Model Dev.* **4** 701–22

DehghaniSanij H, Yamamoto T and Rasiah V 2004 Assessment of evapotranspiration estimation models for use in semi-arid environments *Agric. Water Manage.* **64** 91–106

Deryng D, Conway D, Ramankutty N, Price J and Warren R 2014 Global crop yield response to extreme heat stress under multiple climate change futures *Environ. Res. Lett.* **9** 034011

Dole R, Hoerling M, Perlwitz J, Eischeid J, Pegion P, Zhang T, Quan X-W, Xu T and Murray D 2011 Was there a basis for anticipating the 2010 russian heat wave? *Geophys. Res. Lett.* **38** L06702

Donohue R J, McVicar T R and Roderick M L 2010 Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate *J. Hydrol.* **386** 186–97

Douglas E M, Jacobs J M, Sumner D M and Ray R L 2009 A comparison of models for estimating potential evapotranspiration for Florida land cover types *J. Hydrol.* **373** 366–76

Drewniak B, Song J, Prell J, Kotamarthi V R and Jacob R 2013 Modeling agriculture in the community land model *Geosci. Model Dev.* **6** 495–515

Dury M, Hambuckers A, Warnant P, Henrot A, Favre E, Ouberdous M and François L 2011 Responses of European forest ecosystems to 21st century climate: assessing changes in interannual variability and fire intensity *iForest* **4** 82

Eisner S 2016 Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients *PhD thesis* (Kassel: University of Kassel)

Elliott J, Kelly D, Chryssanthacopoulos J, Glotter M, Jhunjhnuwala K, Best N, Wilde M and Foster I 2014 The parallel system for integrating impact models and sectors (pSIMS) *Environ. Modell. Softw.* **62** 509–16

Elliott J *et al* 2015 The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0) *Geosci. Model Dev.* **8** 261–77

Fader M, Rost S, Müller C, Bondeau A and Gerten D 2010 Virtual water content of temperate cereals and maize: Present and potential future patterns *J. Hydrol.* **384** 218–31

Fisher J B *et al* 2017 The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources *Water Resour. Res.* **53** 2618–26

Fisher J B, Tu K P and Baldocchi D D 2008 Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites *Remote Sens. Environ.* **112** 901–19

Folberth C, Yang H, Wang X and Abbaspour K C 2012 Impact of input data resolution and extent of harvested areas on crop yield estimates in large-scale agricultural modeling for maize in the USA *Ecol. Model.* **235** 8–18

Frieler K *et al* 2015 A framework for the cross-sectoral integration of multi-model impact projections: land use decisions under climate impacts uncertainties *Earth Syst. Dyn.* **6** 447–60

García-Herrera R, Díaz J, Trigo R M, Luterbacher J and Fischer E M 2010 A review of the European summer heat wave of 2003 *Crit. Rev. Environ. Contr.* **40** 267–306

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y and Tanaka K 2008 An integrated model for the assessment of global water resources—Part 1: model description and input meteorological forcing *Hydrol. Earth Syst. Sci.* **12** 1007–25

Hauser M, Orth R and Seneviratne S I 2016 Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia *Geophys. Res. Lett.* **43** 2016GL068036

Henderson-Sellers A, McGuffie K and Pitman A J 1996 The project for intercomparison of land-surface parametrization schemes (PILPS): 1992–1995 *Clim. Dyn.* **12** 849–59

Henderson-Sellers A, Yang Z-L and Dickinson R E 1993 The Project for intercomparison of land-surface parameterization schemes *Bull. Am. Meteorol. Soc.* **74** 1335–49

Hoekstra A 1998 *Perspectives on Water: An Integrated Model-Based Exploration of the Future* (Utrecht: International Books)

Hoerling M, Eischeid J, Kumar A, Leung R, Mariotti A, Mo K, Schubert S and Seager R 2013 Causes and predictability of the 2012 great plains drought *Bull. Am. Meteorol. Soc.* **95** 269–82

Ito A and Inatomi M 2012 Use of a process-based model for assessing the methane budgets of global terrestrial ecosystems and evaluation of uncertainty *Biogeosciences* **9** 759–73

Izaurralde R, Williams J, McGill W, Rosenberg N and Jakas M Q 2006 Simulating soil C dynamics with epic: model description and testing against long-term data *Ecol. Model.* **192** 362–84

Jung M, Reichstein M and Bondeau A 2009 Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model *Biogeosciences* **6** 2001–13

Kay A L, Bell V A, Blyth E M, Crooks S M, Davies H N and Reynard N S 2013 A hydrological perspective on evaporation: historical trends and future projections in Britain *J. Water Clim. Change* **4** 193–208

Kingston D G, Todd M C, Taylor R G, Thompson J R and Arnell N W 2009 Uncertainty in the estimation of potential evapotranspiration under climate change *Geophys. Res. Lett.* **36** L20403

Kiniry J R, Williams J R, Major D J, Izaurralde R C, Gassman P W, Morrison M, Bergentine R and Zentner R P 1995 EPIC model parameters for cereal, oilseed, and forage crops in the northern Great Plains region *Can. J. Plant Sci.* **75** 679–88

Kite G W and Droogers P 2000 Comparing evapotranspiration estimates from satellites, hydrological models and field data *J. Hydrol.* **229** 3–18

Krinner G, Viovy N, de Noblet-Ducoudré N, Ogée J, Polcher J, Friedlingstein P, Ciais P, Sitch S and Prentice I C 2005 A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system *Glob. Biogeochem. Cycles* **19** GB1015

Leng G, Huang M, Tang Q and Leung L R 2015 A modeling study of irrigation effects on global surface water and groundwater resources under a changing climate *J. Adv. Model. Earth Syst.* **7** 1285–304

Liang X, Lettenmaier D P, Wood E F and Burges S J 1994 A simple hydrologically based model of land surface water and energy fluxes for general circulation models *J. Geophys. Res. Atmos.* **99** 14415–28

Lindeskog M, Arneth A, Bondeau A, Waha K, Seaquist J, Olin S and Smith B 2013 Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa *Earth Syst. Dyn.* **4** 385–407

Liu J, Williams J R, Zehnder A J B and Yang H 2007 GEPIC—modelling wheat yield and crop water productivity with high resolution on a global scale *Agric. Syst.* **94** 478–93

Liu J and Yang H 2010 Spatially explicit assessment of global consumptive water uses in cropland: green and blue water *J. Hydrol.* **384** 187–97

Liu W, Yang H, Folberth C, Wang X, Luo Q and Schulin R 2016 Global investigation of impacts of PET methods on simulating crop-water relations for maize *Agric. Forest Meteorol.* **221** 164–75

Martens B, Miralles D G, Lievens H, van der Schalie R, de Jeu R A M, Férnandez-Prieto D, Beck H E, Dorigo W A and Verhoest N E C 2016 GLEAM v3: satellite-based land evaporation and root-zone soil moisture *Geosci. Model Dev. Discuss.* **2016** 1–36

Müller Schmied H *et al* 2016 Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use *Hydrol. Earth Syst. Sci.* **20** 2877–98

Mu Q, Zhao M and Running S W 2011 Improvements to a MODIS global terrestrial evapotranspiration algorithm *Remote Sens. Environ.* **115** 1781–800

Mueller B *et al* 2013 Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis *Hydrol. Earth Syst. Sci.* **17** 3707–20

Mueller B *et al* 2011 Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations *Geophys. Res. Lett.* **38** L06402

Oki T and Sud Y C 1998 Design of Total Runoff Integrating Pathways (TRIP)–A global river channel network *Earth Interact.* **2** 1–37

Ortega-Farias S, Olioso A, Antonioletti R and Brisson N 2004 Evaluation of the Penman-Monteith model for estimating soybean evapotranspiration *Irrigation Sci.* **23** 1–9

Orth R and Seneviratne S I 2015 Introduction of a simple-model-based land surface dataset for Europe *Environ. Res. Lett.* **10** 044012

Pielke R A, Doesken N, Bliss O, Green T, Chaffin C, Salas J D, Woodhouse C A, Lukas J J and Wolter K 2005 Drought 2002 in Colorado: an unprecedented drought or a routine drought? *Pure Appl. Geophys.* **162** 1455–79

Pitman A J 2003 The evolution of, and revolution in, land surface schemes designed for climate models *Int. J. Climatol.* **23** 479–510

Pitman A J, Slater A G, Desborough C E and Zhao M 1999 Uncertainty in the simulation of runoff due to the parameterization of frozen soil moisture using the global soil wetness project methodology *J. Geophys. Res.* **104** 16 879 88

Pokhrel Y N, Koirala S, Yeh P J-F, Hanasaki N, Longuevergne L, Kanae S and Oki T 2015 Incorporation of groundwater pumping in a global Land Surface Model with the representation of human impacts *Water Resour. Res.* **51** 78–96

Portmann F T, Siebert S and Döll P 2010 MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling *Glob. Biogeochem. Cycles* **24** GB1011

Prudhomme C and Williamson J 2013 Derivation of RCM-driven potential evapotranspiration for hydrological climate change impact analysis in Great Britain: a comparison of methods and associated uncertainty in future projections *Hydrol. Earth Syst. Sci.* **17** 1365

Reichle R H, Draper C S, Liu Q, Girotto M, Mahanama S P P, Koster R D and De Lannoy G J M 2017 Assessment of MERRA-2 land surface hydrology estimates *J. Clim.* **30** 2937–60

Reichle R H, Liu Q, Koster R D, Draper C S, Mahanama S P P and Partyka G S 2016 Land surface precipitation in MERRA-2 *J. Clim.* **30** 1643–64

Sacks W J, Deryng D, Foley J A and Ramankutty N 2010 Crop planting dates: an analysis of global patterns *Glob. Ecol. Biogeogr.* **19** 607–20

Sanderson B M and Knutti R 2012 On the interpretation of constrained climate model ensembles *Geophys. Res. Lett.* **39** L16708

Sanderson B M, Knutti R and Caldwell P 2015 Addressing interdependency in a multimodel ensemble by interpolation of model properties *J. Clim.* **28** 5150–70

Schellekens J *et al* 2017 A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset *Earth Syst. Sci. Data* **9** 389–413

Schär C, Vidale P L, Lüthi D, Frei C, Häberli C, Liniger M A and Appenzeller C 2004 The role of increasing temperature variability in European summer heatwaves *Nature* **427** 332–6

Schulz J-P, Dümenil L, Polcher J, Schlosser C A and Xue Y 1998 Land surface energy and moisture fluxes: comparing three models *J. Appl. Meteorol.* **37** 288–307

Schwingshackl C, Hirschi M and Seneviratne S I 2017 Quantifying spatiotemporal variations of soil moisture control on surface energy balance and near-surface air temperature *J. Clim.* **30** 7105–24

Seneviratne S I 2012 Climate science: historical drought trends revisited *Nature* **491** 338–9

Seneviratne S I, Corti T, Davin E L, Hirschi M, Jaeger E B, Lehner I, Orlowsky B and Teuling A J 2010 Investigating soil moisture-climate interactions in a changing climate: a review *Earth Sci. Rev.* **99** 125–61

Shao Y and Henderson-Sellers A 1996 Modeling soil moisture: a Project for Intercomparison of land surface parameterization schemes phase 2(b) *J Geophys. Res. Atmos.* **101** 7227–50

Shaw S B and Riha S J 2011 Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests *Hydrol. Process.* **25** 1466–78

Sheffield J, Goteti G and Wood E F 2006 Development of a 50 year high-resolution global dataset of meteorological forcings for land surface modeling *J. Clim.* **19** 3088–111

Sheffield J, Wood E F and Roderick M L 2012 Little change in global drought over the past 60 years *Nature* **491** 435–8

Sitch S *et al* 2003 Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model *Glob. Change Biol.* **9** 161–85

Smith B, Wårlind D, Arneth A, Hickler T, Leadley P, Siltberg J and Zaehle S 2014 Implications of incorporating n cycling and n limitations on primary production in an individual-based dynamic vegetation model *Biogeosciences* **11** 2027–54

Sperna Weiland F C, Tisseuil C, Dürr H H, Vrac M and van Beek L P H 2012 Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study *Hydrol. Earth Syst. Sci.* **16** 983–1000

Stacke T and Hagemann S 2012 Development and evaluation of a global dynamical wetlands extent scheme *Hydrol. Earth Syst. Sci.* **16** 2915–33

Su Z 2001 The surface energy balance system (SEBS) for estimation of turbulent heat fluxes *Hydrol. Earth Syst. Sci.* **6** 85–100

Tang Q, Oki T, Kanae S and Hu H 2007 The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation *J. Hydrometeorol.* **8** 499–512

Teuling A J, Van Loon A F, Seneviratne S I, Lehner I, Aubinet M, Heinesch B, Bernhofer C, Grünwald T, Prasse H and Spank U 2013 Evapotranspiration amplifies European summer drought *Geophys. Res. Lett.* **40** 2071–5

Tian H, Chen G, Lu C, Xu X, Hayes D J, Ren W, Pan S, Huntzinger D N and Wofsy S C 2015 North American terrestrial $CO_2$ uptake largely offset by $CH_4$ and $N_2O$ emissions: toward a full accounting of the greenhouse gas budget *Clim. Change* **129** 413–26

Trenberth K E, Branstator G W and Arkin P A 1988 Origins of the 1988 North American Drought *Science* **242** 1640–5

van Dijk A I J M, Peña-Arancibia J L, Wood E F, Sheffield J and Beck H E 2013 Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide *Water Resour. Res.* **49** 2729–46

Vinukollu R K, Meynadier R, Sheffield J and Wood E F 2011 Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends *Hydrol. Process.* **25** 3993–4010

Vörösmarty C J, Federer C A and Schloss A L 1998 Potential evaporation functions compared on US watersheds: possible implications for global-scale water balance and terrestrial ecosystem modeling *J. Hydrol.* **207** 147–69

Wada Y, Wisser D and Bierkens M F P 2014 Global modeling of withdrawal, allocation and consumptive use of surfacewater and groundwater resources *Earth Syst. Dyn.* **5** 15–40

Wang K, Dickinson R E, Wild M and Liang S 2010a Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development *J. Geophys. Res. Atmos.* **115** D20112

Wang K, Dickinson R E, Wild M and Liang S 2010b Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 2. results *J Geophys. Res. Atmos.* **115** D20113

Ward J H 1963 Hierarchical grouping to optimize an objective function *J. Am. Stat. Assoc.* **58** 236–44

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 The inter-sectoral impact model intercomparison project (ISI MIP): project framework *Proc. Natl Acad. Sci.* **111** 3228–32

Weedon G P, Balsamo G, Bellouin N, Gomes S, Best M J and Viterbo P 2014 The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data *Water Resour. Res.* **50** 7505–14

Weedon G P, Gomes S, Viterbo P, Shuttleworth W J, Blyth E, Österle H, Adam J C, Bellouin N, Boucher O and Best M 2011 Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century *J. Hydrometeorol.* **12** 823–48

Weiß M and Menzel L 2008 A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments *Adv. Geosci.* **18** 15–23

Williams J R 1995 The epic model *Computer Models of Watershed Hydrology* ed V P Singh (Highlands Ranch, CO: Water Resources) pp 909–1000

Wood E F *et al* 1998 The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c) Red–Arkansas river basin experiment: 1. Experiment description and summary intercomparisons *Glob. Planet. Change* **19** 115–35

Wu X *et al* 2016 ORCHIDEE-CROP (v0), a new process-based agro-land surface model: model description and evaluation over Europe *Geosci. Model Dev.* **9** 857–73

Xu C-Y and Chen D 2005 Comparison of seven models for estimation of evapotranspiration and groundwater recharge using lysimeter measurement data in Germany *Hydrol. Process.* **19** 3717–34

Zeng N, Mariotti A and Wetzel P 2005 Terrestrial mechanisms of interannual $CO_2$ variability *Glob. Biogeochem. Cycles* **19** GB1016

Zeng Z, Wang T, Zhou F, Ciais P, Mao J, Shi X and Piao S 2014 A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982–2009 *J. Geophys. Res. Atmos.* **119** 2013JD020941

Zhang Y *et al* 2016 Multi-decadal trends in global terrestrial evapotranspiration and its components *Sci. Rep. UK* **6** 19124