A global scale evaluation of extreme event uncertainty in the *eartH2Observe* project

Toby R. Marthews[1], Eleanor M. Blyth[1], Alberto Martínez-de la Torre[1], Ted I. E. Veldkamp[2]

[1]Centre for Ecology & Hydrology, Maclean Building, Wallingford OX10 8BB, U.K.
[2]Institute for Environmental Studies, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

Correspondence to: Toby R. Marthews (tobmar@ceh.ac.uk)

Abstract. Knowledge of how uncertainty propagates through a hydrological land surface modelling sequence is of crucial importance in the identification and characterisation of system weaknesses in the prediction of droughts and floods at global scale. We evaluated the performance of five state-of-the-art global hydrological and land surface models in the context of modelling extreme conditions (drought and flood). Uncertainty was apportioned between model used (model skill) and also the satellite-based precipitation products used to drive the simulations (forcing data variability) for extreme values of precipitation, surface runoff and evaporation. We found in general that model simulations acted to augment uncertainty rather than reduce it. In percentage terms, the increase in uncertainty was most often less than the magnitude of the input data uncertainty, but of comparable magnitude in many environments. Uncertainty in predictions of evapotranspiration lows (drought) in dry environments was especially high, indicating that these circumstances are a weak point in current modelling system approaches. We also found that high data and model uncertainty points for both ET lows and runoff lows were disproportionately concentrated in the equatorial and southern tropics. Our results are important for highlighting the relative robustness of satellite products in the context of land surface simulations of extreme events and identifying areas where improvements may be made in the consistency of simulation models.

1 Introduction

Producing robust predictions about the future dynamics of the water cycle at local, regional and global scales is critically important because it is the only way to avoid or mitigate the effects of water cycle extremes (e.g. flood, drought) (IPCC, 2012) and, in the longer term, to improve our use of resources and achieve long-term adaptation to climate change (Bierkens, 2015). Over the 21st century, climate and hydrological regimes are predicted to undergo significant shifts in baseline variables such as temperature, precipitation and runoff, leading to changes in the frequency of extremes of precipitation, evaporation and overland flow, and ultimately to changes in the frequency and intensity of both floods and droughts (Bierkens, 2015; Dadson et al., 2017; Prudhomme et al., 2014). Understanding and predicting these shifts in the global dynamical system, both at atmospheric and land surface level is therefore of crucial importance (Santanello *et al.* 2018).

All model predictions have uncertainties, and linked modelling sequences have identifiable uncertainties at each step in the sequence (uncertainty propagation). In the case of a hydrological land surface modelling sequence, where climate data inputs are used to drive a simulator of the surface water cycle and land surface interactions, there are two main sources of uncertainty: *data uncertainty* (differences between forcing data used) and *model uncertainty* (differences between the simulation models). Data and model uncertainty differ greatly not just between themselves at particular locations, but also between coastal and floodplain areas of the world, and remote regions with heterogeneous terrain (Bhuiyan et al., 2018a; Riley et al., 2017) and between extreme high flows (floods) (Mehran and AghaKouchak, 2014; Nikolopoulos et al., 2016) and extreme water scarcity (droughts) (Veldkamp and Ward, 2015).

We focus on the relative dominance of model uncertainty (we take this as a broadly defined measure, including uncertainty from hydrology models that simulate water dynamics, vegetation models that focus on carbon dynamics and land surface models that attempt to integrate all biogeochemical cycles) and uncertainty in the precipitation product used to drive those models. In situations where model uncertainty is significant, the range of predictions possible from standard model simulations is

of great importance to stakeholders and other users. If precipitation data uncertainty dominates, however, then greater attention should arguably be focused on selecting the most appropriate product to use, and perhaps additionally on interrogating the potentially sparse data base of precipitation measuring stations used by the precipitation products.

1.1 Uncertainties in land surface model simulations

Model uncertainty, i.e. prediction variation as a result of differing process representations within a model (e.g. Li and Wu (2006)), is commonly the dominant uncertainty in complex systems used in risk-informed decision-making (Oberkampf and Roy, 2010). Although historically often overlooked (Li and Wu, 2006), model uncertainty has recently come under increasing scrutiny in the context of land surface models (Huntingford et al., 2013; Long et al., 2014; Schewe et al., 2014; Ukkola et al., 2016). A lack of adequate representation of flood-generation processes (both from surface and subsurface runoff) and permafrost or snow dynamics can lead to an imprecise simulation of runoff peaks in many large river basins, and a lack of proper representation of wetland evaporation and human effects such as water consumption and inter-basin transfers can lead to over- or under-estimated discharge in many basins, especially those with large semiarid regions (Bierkens, 2015; Veldkamp et al., 2018). Additionally, even though regional-scale precipitation is predominantly caused by the atmospheric moisture convergence associated with large-scale and mesoscale circulations, processes operating on smaller length scales significantly modify even regional-scale dynamics, so it is to be expected that uncertainty in land surface models will depend on local topography, the presence or absence of vegetation or water bodies and, importantly, which type of precipitation is dominant at a particular point and time (cyclonic, orographic or convective, Table 1).

1.2 Uncertainties in precipitation products

Precipitation is a necessary forcing input for land surface and hydrological models that is extremely challenging to estimate independently (Beck et al., 2017b; Bhuiyan et al., 2018a; Bhuiyan et al., 2018b; Levizzani et al., 2018). The accuracy and precision of precipitation measurements fundamentally influences predictions of land surface and hydrological models (Hirpa et al., 2016), however many widely-used precipitation products have high uncertainties over the tropics and/or areas of high relief (Bierkens, 2015; Derin et al., 2016; Kimani et al., 2017; Yin et al., 2015).

High precipitation extremes are not always well-characterised: Mehran and AghaKouchak (2014) reviewed the capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations. For example, the precipitation radar on board TRMM (Table 2) is capable of capturing moderate to heavy precipitation but does not detect light rain or drizzle (Huffman et al., 2007; Luo et al., 2017).

Low precipitation extremes are also not always well-characterised: Veldkamp and Ward (2015) reviewed the advantages of different drought indices and highlighted many issues at the global scale. This relates to a more general point about remote sensing rainfall intensity: a precipitation product is more likely to record correctly that it is raining at a particular location than to record correctly the amount, which is unfortunate because it is usually precipitation amount that is most important for predictive modelling of drought or flood intensity.

Accuracy of meteorological data including precipitation will be expected to be lower (and uncertainty higher) for 'real-time' precipitation products because they have not been 'blended' with raingauge or reanalysis data (Table 2) (Munier et al., 2018). If a near-real time estimate of drought or flood is needed, therefore, then a cost-benefit balance arises with the end user having to make a choice between up-to-date information versus lowest uncertainty (Munier et al., 2018).

1.3 The *eartH2Observe* project

During 2014-2018, the *eartH2Observe* project http://www.eartH2Observe.eu/ brought together a multinational team of modelling and Earth Observation (EO) researchers to improve the assessment of global water resources through the integration of new datasets and modelling techniques. The uncertainties described above for different parts of the forcing data - land surface model system have been the starting point for this investigation, and *eartH2Observe* has quantified these uncertainties using an ensemble

of forcing data and modelling systems. The project aimed to provide an overall understanding of the uncertainty in the EO products and EO-driven water resources models. This understanding is needed for optimal data-model integration and for water resources reanalysis, and their use for basin scale and end-user applications (e.g. floods, droughts, basin water budgets, stream flow simulations) (Nikolopoulos et al., 2016). As part of *eartH2Observe*, and in order to make progress towards this aim, in this study we asked the following two research questions:

(1) Under what circumstances can uncertainty in the prediction of water cycle quantities be attributed clearly to the model in use (model uncertainty) and/or to the precipitation product used to drive the model (data uncertainty)?

(2) When uncertainty is attributable to both model and data sources, is data uncertainty generally the greater (i.e. the model acts to reduce or 'stabilise' uncertainty) or the lesser (i.e. the model effectively augments variation)?

2 Data and methods

Uncertainty in extreme event representation varies both between models used (model uncertainty) and also between satellite-based precipitation products used to drive the simulations (data uncertainty). Five of the most widely-used and well-supported precipitation data products were used in this study (Table 2) and five state-of-the-art land surface models and hydrological models were run using each of those forcing data products (Table 3). This produced an ensemble of 25 estimates for each output variable (mirroring the *method of competing models* approach advocated for complex systems by (Oberkampf and Roy, 2010)).

Only the precipitation forcing data for each model were allowed to vary between simulations: the remaining non-precipitation drivers (temperature, wind speed, radiation, etc.) were held constant across all simulations and taken from global Water Resources Reanalysis 2 baseline forcing data used in other *eartH2Observe* projects (WRR2) (Arduini et al., 2017). The combination of WRR2 non-precipitation drivers and the selected precipitation drivers (Table 2) is called WRR-ENSEMBLE (Arduini et al., 2017). All simulations used a global spatial resolution of 0.25° and covered the period 2000-2013. Because of source data limitations (Table 2), we restricted our analysis to latitudinal zones between 50°S and 50°N (Fig. 1).

2.1 Focus on extremes

Performance was assessed in terms of the variability of evapotranspiration (ET) and surface runoff under extreme rainfall conditions (both high extremes and low extremes). We quantified the relative magnitudes of these uncertainties under (i) varying simulation model (model uncertainty) and (ii) varying choice of precipitation product (data uncertainty). We quantified uncertainty in terms of the number of extreme events per month, with the *extreme event* defined as the occurrence of an extreme value for the monthly average of a given variable, and *extreme* defined as a value in the top/bottom 10% of the baseline distribution of values for that variable (IPCC, 2014). Extreme event probability was calculated within each pixel for each month of the year, summed over the year and then the standard deviation (SD) taken across either the model outputs or precipitation products (see below) in units of (occurrence of extreme events per year). In order to avoid spurious extremes occurring in deserts and other areas with very low variability in water cycle values, gridcells with less than 20 mm annual precipitation (multi-year mean) or <0.1 SD in their monthly precipitation across the year were excluded.

Extremes for any particular variable may only be assessed in relation to an estimate of 'normal' conditions, and for this we took a baseline distribution of values calculated at each gridcell (i.e. not globally, regionally or per biome) from an average of the five simulations involving the 2000-2013 MSWEP forcing data (Beck et al., 2017a). We took MSWEP to be our 'gold standard' product in this sense because of its high reliability and multi-source nature (satellite observations blended with reanalysis and gauge data, Beck et al. (2017a), Munier et al. (2018)) in comparison to other available products (Table 2).

2.2 Uncertainty propagation

We defined three indices of uncertainty propagation α, β and ε to quantify the extent to which a given land surface or hydrological model increases or augments the uncertainty introduced to its simulations via the precipitation driver inputs (we consider water cycle variables only in this analysis so it is reasonable to assume that uncertainty in our variables is not independent of uncertainty in precipitation). The α measure quantifies the increase or decrease in uncertainty attributable to the precipitation drivers, β measures the equivalent for uncertainty attributable to the simulator model itself and ε quantifies the overall change in uncertainty over the course of the simulation. Note that the quantification of absolute uncertainty in predicted quantities (Li and Wu, 2006) is not our focus: we are instead concerned with the relative contributions of data and model uncertainty in a combination setting (Oberkampf and Roy, 2010).

The defining equations are (calculated on a gridcell by gridcell basis):

$$\text{Scaled data uncertainty } \alpha_{X,j} = DOU \div DIU \quad (1)$$

$$\text{Scaled model uncertainty } \beta_{X,j} = MU \div DIU \quad (2)$$

$$\text{Scaled total uncertainty } \varepsilon_{X,j} = \alpha_{X,j} + \beta_{X,j} = (DOU + MU) \div DIU \quad (3)$$

where *DIU* = Mean uncertainty across products in precipitation extreme occurrence (input forcing data uncertainty)

DOU = Mean uncertainty across products in variable *X* extreme occurrence (output model uncertainty attributable to forcing data input)

MU = Mean uncertainty across models in variable *X* extreme occurrence (output model uncertainty attributable to model differences)

All mean uncertainties are in units of (extreme event occurrence frequency per year: EE/yr hereafter) and *j* can be either *high* or *low* depending on whether high or low extremes are being considered. The uncertainty propagation involves input uncertainty from the precipitation driver (*DIU*), which under the simulation is modified into the uncertainty of *X* when averaged across the different results obtained from using different precipitation products (*DOU*), but, unlike the forcing data, the simulation results have uncertainty as a consequence of the differences between simulator model used (*MU*) which means that total uncertainty at output level is (*DOU*+*MU*).

In summary, $\varepsilon_{X,j}$ may be understood as a measure of how much input precipitation product data uncertainty (*DIU*) is amplified into output uncertainty (*DOU*+*MU*) during an ensemble of simulations. $\alpha_{X,j}$ may be understood as the special case of $\varepsilon_{X,j}$ where the ensemble consists of one model only, and $\beta_{X,j}$ as the special case of $\varepsilon_{X,j}$ where all ensemble members use one precipitation product only. Values of $\varepsilon_{X,j} > 1.0$ indicate that the model simulation acts effectively to increase (amplify) the uncertainty in the forcing precipitation data. Values $0.0 < \varepsilon_{X,j} < 1.0$ indicate that the model simulation acts to decrease (stabilise) the uncertainty in the forcing precipitation data (i.e. the model acts effectively to 'stabilise' the input uncertainty to ($\varepsilon_{X,j}$*100)% of the input data uncertainty). Division by zero should not occur because of the masking to avoid spurious 'extremes' in arid areas (above).

3 Results

Comparison of precipitation extreme event occurrences across the forcing precipitation products shows immediate differences both spatially (Fig. 2) and between the products themselves (Fig. 3). Notably, the precipitation products differ in their extreme event occurrence rates, with especially TRMM-RT presenting increased rates of extreme high precipitation events across the globe and particularly GSMaP presenting increased rates of extreme low events (for uncertainty maps, see Fig. S1, Fig. S2, Fig. S3 and Fig. S4). Calculating these absolute uncertainty values is a necessary step towards assessing the relative magnitudes of data and model uncertainty for different extreme events.

3.1 Scaled uncertainty

Considering firstly $\alpha_{X,j}$, the uncertainty that is directly attributable to the precipitation data products, we found that in terms of global average $\alpha_{X,j}$ was mostly <1 (i.e. $\log_{10}(\alpha_{X,j})<0$) for ET highs (58.1% *vs.* 41.9%) and decreased as precipitation increased in all latitudinal zones except the northern tropics, but for runoff highs, $\alpha_{X,j}$ increased with precipitation in all latitudinal zones except the equatorial tropics (Fig. 4). Points where data uncertainty greatly increased on propagation through models ($\alpha_{X,j}>1$) occurred mostly during the prediction of low extremes (ET or runoff) and were restricted to areas with rainfall <2000 mm/yr (Fig. 4). Points where data uncertainty greatly decreased on propagation through models ($\alpha_{X,j}<0.1$, $\log_{10}(\alpha_{X,j})<-1$) occurred mostly during the prediction of runoff extremes (mostly low extremes, but also high) and were restricted to areas with rainfall <1000 mm/yr (Fig. 4). Points with high precipitation uncertainty occurred in both dry and wet environments.

Considering $\beta_{X,j}$, the increase in model uncertainty relative to input data uncertainty, we found that $\beta_{X,j}$ was dominantly <1 (i.e. $\log_{10}(\beta_{X,j})<0$) for ET highs (80.1% *vs.* 19.8%) and decreased as precipitation increased in all latitudinal zones; for runoff highs, $\beta_{X,j}$ was also mostly <1 (55.6% *vs.* 44.4%) but increased with precipitation in all latitudinal zones except the equatorial tropics (Fig. 5).

The scaled increase in total (data + model) uncertainty is measured by $\varepsilon_{X,j}$. In all latitude zones except the northern tropics, we found that uncertainty in ET highs increased over the course of the simulation ($\varepsilon_{X,j}$ was dominantly >1 - i.e. $\log_{10}(\varepsilon_{X,j})>0$) at the great majority of locations (80.5% *vs.* 19.5%), though the magnitude of the increase reduced in wetter environments (Fig. 6). In all latitude zones except the equatorial tropics, we also found that uncertainty in runoff highs increased over the course of the simulation at the great majority of locations (76.2% *vs.* 23.8%), but for runoff the magnitude increased with precipitation (Fig. 6). This implies that the causes of higher model uncertainty operate differentially in wet and dry environments, with dry environments being perhaps generally less well-modelled than wetter environments.

3.2 Global uncertainty

The global mean value of α is a measure of the amount a given quantity is affected as precipitation changes relative to the input precipitation data uncertainty (Eq. 1). For quantities that 'track precipitation' (i.e. are sensitive to precipitation extremes), we would expect this to be close to 1 (e.g. runoff values, Fig. 7a), but especially in drier climates small variations in precipitation can drive much higher variation in output variables through threshold effects, so we might expect higher values in such regions (e.g. ET values, Fig. 7b).

The global mean value of β_X is a measure of the internal model uncertainty in quantity X, relative to the input precipitation data uncertainty (Eq. 2), i.e. a measure of the diversity of the calculation methods used to derive X between models. If quantity X is sensitive to precipitation extremes, we should expect low model uncertainty and therefore low values of β_X (e.g. under conditions where evapotranspiration and soil storage are minimal we would expect runoff highs and lows to be closely similar to precipitation highs and lows with the model introducing little modification of the input data). Our results show that evapotranspiration extremes are more sensitive to precipitation uncertainty in wet environments than dry environments (Fig. 7c).

Globally, model uncertainty was generally less than data uncertainty (Fig. 5, Fig. 7). In the equatorial tropics, ET prediction uncertainty was more attributable to data uncertainty, but runoff uncertainty was more attributable to model uncertainty, either indicating a wider variety of model representations of runoff generation processes within the tested models, or a greater dependence of ET estimates on precipitation inputs (Fig. 5).

Munier et al. (2018) found that the occurrence of flood (high runoff values) is generally more sensitive to high precipitation extremes than the occurrence of high evapotranspiration values, but that the reverse is true for low extremes. We do find this in our results as a rule of thumb across all environments (e.g. ($\varepsilon_{ET,high}<\varepsilon_{runoff,high}$) and ($\varepsilon_{ET,low}>\varepsilon_{runoff,low}$) and the same for α and β in Fig. 7a), but we also note that in very dry and very wet environments this pattern does not persist (Fig. 7) and it also does not persist in all latitudinal zones when taken separately.

The total change in uncertainty over the course of the simulation of variable X is measured by $\varepsilon_{X,j}$ (Eq. 3) and our values for $\varepsilon_{X,j}$ were universally >1.0, indicating that the model simulation does act effectively to increase (amplify) the uncertainty in the

forcing precipitation data. This also implies that when a set of models is under consideration, model uncertainty is usually greater than data uncertainty. Finally, high uncertainty points for ET lows and runoff lows were disproportionately concentrated in the equatorial and southern tropics not only for $\varepsilon_{X,j}$ but also for both components $\alpha_{X,j}$ and $\beta_{X,j}$ (Fig. 4, Fig. 5 and Fig. 6; cf. Fig. 2).

4 Discussion

Model output uncertainty is always a mixture of input data uncertainty and uncertainty accumulated during the simulation (Li and Wu, 2006; Oberkampf and Roy, 2010; Van Loon, 2015). However, these uncertainties are unfortunately not orthogonal in general because the models encode nonlinear relationships and therefore cannot be assumed to react consistently to different levels of precipitation input (e.g. (Bhuiyan et al., 2018a; Munier et al., 2018; Ukkola et al., 2016)). In this study we have had unprecedented access through the *eartH2Observe* project to an ensemble of simulations that has combined a selection of widely-used and validated precipitation data products with a spread of cutting edge land surface and hydrology simulation models.

4.1 Clear attribution of uncertainty to data and/or model sources

Under what circumstances can uncertainty in the prediction of water cycle quantities be attributed clearly to the model in use (model uncertainty) and/or to the precipitation product used to drive the model (data uncertainty)? Ukkola et al. (2016) found that land surface models diverged in evapotranspiration prediction during the dry season, and the results of our study strongly support this conclusion, with our calculated envelope of uncertainty widening in drier climates across the globe for all our uncertainty measures.

We found that high data and model uncertainty points for both ET lows and runoff lows were disproportionately concentrated in the equatorial and southern tropics. These zones are dominantly covered by tropical rainforests and savanna grasslands, so one possibility is that low fluxes in xeric environments are better characterised - both in data products and model characterisation - than low fluxes in these mesic and hydric environments. Data products are known to be more accurate away from areas with consistent cloud cover and a high occurrence of convective rainfall (Table 1) (Derin et al., 2016; Levizzani et al., 2018), which might explain this for data uncertainty, but having model uncertainty follow the same geographic distribution indicates that we must also consider uncertainties in the calculations of runoff and evapotranspiration. It seems also to be the case that the simple water balance approach taken by land surface and hydrology models becomes approximate in latitudinal zones where low flows are generally combined with higher temperatures and more episodic rainfall events (McGregor and Nieuwolt, 1998). This could indicate that using generalised approaches for all environments (e.g. the Priestley-Taylor or Penman-Monteith equations) is no longer sufficient for simulations at these spatio-temporal scales (Long et al., 2014; Wartenburger et al., 2018) or perhaps because we still lack crucial processes in these models, e.g. soil crusting or sealing, which only occur in semi-arid or arid areas (Marshall et al., 1996). However, we must also be careful to draw strong conclusions from these zones because another possibility is that this result simply confirms that these regions are where our available sources data are of lower quality (q.v. Fig. 2a).

Uncertainty in predictions of evapotranspiration lows (drought) in dry environments is especially high, indicating that these circumstances are a weak point in current modelling approaches. Importantly, our results quantify this effect and show that even though uncertainty in the precipitation inputs is highest in these environments, the uncertainty in model representation of the processes involved is also significant and should not be ignored. A practical application of this is that when robust predictions of drought are required in very dry environments, not only should a spread of precipitation products be applied, but also more than one simulator model, and the model outputs should be validated as closely as possible against local data sources in order to ensure that conclusions drawn from these analyses are suitable for decision-making.

4.2 Relative importance of data and model uncertainty

When uncertainty is attributable to both model and data sources in a simulation ensemble, is data uncertainty generally the greater (i.e. the model acts to reduce or ‘stabilise’ uncertainty) or the lesser (i.e. the model effectively augments variation)? In a report for

the Intergovernmental Panel on Climate Change (IPCC), Bates et al. (2008) drew attention to the high uncertainty there was in climate models in precipitation data (= *data uncertainty*), and also suggested that for aspects of the hydrological cycle such as changes in evaporation, soil moisture and runoff, the relative spread in projections (= *total uncertainty*) was similar to, or larger than, the changes in precipitation (points echoed later by Schewe et al. (2014) and others). Precipitation observations are known to have high uncertainty (Beck et al., 2017a; Bierkens, 2015; Kimani et al., 2017; Levizzani et al., 2018; Yin et al., 2015), but responses to precipitation low extremes (drought) should not be expected to be proportional to responses from the same model to precipitation high extremes (flood) (Veldkamp et al., 2018).

We found in general that the model simulations we analysed acted to augment uncertainty rather than reduce it. In percentage terms, the increase in uncertainty was most often less than the magnitude of the input data uncertainty, but uncertainty did not decrease through the model for any variable so the simulation models did not in any case act to 'stabilise' or decrease the uncertainty supplied to them through the precipitation data products used to drive them. We do agree with Wartenburger et al. (2018)'s finding that the forcing (data uncertainty) generally dominates the variance in ET extremes, but we found model uncertainty to be important in all cases analysed and very nearly the magnitude of the forcing uncertainty in both very dry and very wet environments. This is a very significant result because it implies that a focus on the reduction of both data and model uncertainty will be necessary in order to improve the prediction of water cycle extremes.

4.3 Sources of unquantified uncertainty

It is important to bear in mind that some sources of uncertainty exist in these water cycle quantities that are as yet unmeasured in any existing data products, and therefore cannot be analysed in this study. There is a very strong current emphasis in climate science on identifying global areas of high precipitation uncertainty, for example (Bierkens, 2015; He et al., 2017; Levizzani et al., 2018), from which we can highlight two: Firstly, most precipitation products record observations of amount, not the type of precipitation (Table 2), however it is very likely that precipitation type strongly influences our precipitation data uncertainty: for example, convective processes are dominant in the precipitation generating processes in dryland ecosystems (Table 1), and different precipitation types occur at different spatial scales as well (Table 1). Secondly, our equatorial tropical zone (Fig. 1) includes the Inter-Tropical Convergence Zone (ITCZ) of low pressure, characterised by convective activity generating many storms. It is well-known that because of the transitory nature of the cloud dynamics in the ITCZ, precipitation products necessarily have higher uncertainty and, simultaneously, these conditions are of too short duration to be captured reliably in our analysis.

For evapotranspiration in particular, Lopez et al. (2017) drew attention to the global lack of high quality *in situ* site data and the "inevitable scale mismatch" when using such data to calibrate Earth Observation datasets. Regional estimates of evapotranspiration rely on scaling-up methods to take account of regional advection effects and, additionally, the use of estimated values for evaporation rates from unmeasured land use types. Each step in these calculations potentially introduces significant uncertainty with the result that there is currently wide variation between the values suggested by even the best global evapotranspiration products (Martens et al., 2017).

Finally, runoff: Surface runoff estimates are linked to precipitation and evapotranspiration estimates via the water cycle balance equation (Beck et al., 2017b; Bierkens, 2015; Veldkamp et al., 2018). Because soil storage terms are usually taken as constant, underestimation of evapotranspiration often means overestimation of runoff and streamflow data (and *vice versa*). In this way, uncertainty in surface runoff is related to uncertainty in evapotranspiration estimates. However, because of the wide availability and high quality of global streamflow datasets (e.g. the Global Runoff Database, GRDC), and a much lower requirement for approximation and gap-filling in comparison to evapotranspiration data, runoff data is usually considered to be of the highest quality in water balance studies.

4.4 Conclusions

Water resources management has become one of the most important challenges facing hydrologists and decision-makers at state and national levels, motivated by increasing water scarcity in some global regions and a higher frequency of extreme flood events

in others (Bierkens, 2015; Dadson et al., 2017; Schewe et al., 2014). At the same time, precipitation extremes are predicted to increase in frequency and impact under committed climate change (Ali and Mishra, 2017). Therefore, reliance on robust model predictions has never been greater (Kundzewicz and Stakhiv, 2010; Riley et al., 2017). In this study we have used an ensemble of simulation results from the *eartH2Observe* project derived from cutting-edge model simulators driven by the best available published (and validated) precipitation observations, but the sources of uncertainty are nevertheless many and varied.

We found that models always augmented uncertainty relative to the magnitude of forcing data uncertainty. Although, for predicting the extremes of evapotranspiration and runoff, the uncertainties inherent in the current generation of precipitation observation products are generally larger than the uncertainty introduced into the calculation by the land surface and hydrology models used, model uncertainty cannot be ignored and in many environments is comparable in magnitude to forcing data uncertainty. Therefore, in order to reduce prediction uncertainty we need very much to make progress on two fronts: (1) we need precipitation data product uncertainty to be reduced (improved satellites are always welcome, of course, but we believe that much progress can also be made through moving towards blended products that are sensitive to more types of precipitation) and (2) we need to improve the mechanistic equations used in these models to derive water cycle quantities (including a better consideration of scale issues and domains of validity for existing equations).

It is important to resolve both data and model uncertainty much more clearly and identify exactly at which points in our linked modelling systems these uncertainties become the most significant. Our current model representation of land surface hydrological and biogeochemical processes remains approximate especially in very dry and very wet environments and there is a clear need for a better characterisation of these environmental extremes in order for us to move forward to the next generation of climate and land surface prediction models.

Table 1: Types of precipitation and their main controlling factors (McGregor and Nieuwolt, 1998).

Precipitation type	Spatial scale	Characteristics	Challenges
Cyclonic (=frontal)	Synoptic, regional	The leading edge of a warm and moist air mass (warm front) meets a cool, dry air mass (cold front). The warmer air mass rises over the cooler air, with precipitation occurring along the front. If the air begins to circulate, a cyclonic storm can occur.	• It is widely accepted that global warming will lead to a higher water-holding capacity for the atmosphere as well as increased rates of evaporation, and therefore increased extreme weather (Trenberth et al., 2015; Yi et al., 2015). However, the mechanisms through which the location and magnitude of these extreme events may be predicted (e.g. tipping points, thresholds) remain inadequately understood (Marthews et al., 2012).
Orographic	Intermediate	Warm, moist air entering a mountain range is forced to rise, and then cools and precipitation ensues (= *orographic lift*).	• Scale is an important issue: mountains can modify large-scale circulation, causing changes in local moisture convergence, but local condensation and microphysical processes also influence flow stability upstream (Marthews et al., 2012).
Convective	Local (often sub-grid)	A warm soil or vegetation surface warms the air above it, which then rises vertically and cools, with precipitation occurring on cooling. 'Convection-permitting' model runs usually require a sub-daily timestep and <10 km spatial resolution, and in the absence of these a convection parameterisation scheme (CPS) is necessary (i.e. assumptions about subgrid and subdaily dynamics) (Prein *et al.* 2015).	• *Stratiform precipitation* is when the rise is diagonal rather than vertical (i.e. similar to orographic, but not as a result of landform) • Sub-grid displacement of cloud occurrence from driver (Taylor et al., 2012) • Land surface exchange (e.g. evapotranspiration) has a significant effect, but often not modelled explicitly. • Resolution of snow versus rainfall in mountain regions is critical for water resources management, but not well-characterised in models. • CPSs generally overestimate light rain (drizzle) because they overestimate the number of precipitation days (by equating clouds with rain) and / or underestimate precipitation intensity (Marthews et al., 2012; Prein et al., 2015). Conversely, it is a known limitation of some satellites that they are not sensitive to, and therefore underestimate, light rain (e.g. Luo et al. (2017)). This introduces a 'calibration gap': calibration of large-scale models against satellite-based precipitation observations must not only factor out the overestimation of CPSs, but also the underestimation of the observations.

Table 2: Global precipitation products used to drive the models selected from Dorigo *et al.* (2014). Data files used are available through the Water Cycle Integrator https://wci.eartH2Observe.eu/ at 25 km resolution for the period 2000-2013. Algorithm type is as given by the International Precipitation Working Group (IPWG) *.

Product	Algorithm	Notes
Multi-Source Weighted-Ensemble Precipitation (MSWEP)		Global reanalysis data (Beck et al., 2017a)
Climate Prediction Center MORPHing Technique (CMORPH)	Blended microwave-infrared	Restricted to 60°S to 60°N A passive microwave-based product advected in time using geosynchronous infrared data (Joyce et al., 2004). When microwave observations are not available, infrared observations are used to advect the last microwave scan over time. In addition to advecting precipitation forward in time, the algorithm propagates precipitation backward once the next microwave observation becomes available (Mehran and AghaKouchak, 2014).
Global Satellite Mapping of Precipitation (GSMaP)	Blended microwave-infrared	Restricted to 60°S to 60°N (Tian et al., 2010)
Tropical Rainfall Measuring Mission (TRMM)	Satellite-based	Restricted to 50°S to 50°N
TRMM Real Time (TRMM-RT)	Satellite-based	Restricted to 50°S to 50°N Mainly based on microwave data aboard Low Earth Orbit satellites (Huffman et al., 2007). The TRMM-RT algorithm is primarily based on microwave observations from low orbiter satellites. Gaps in microwave observations are filled with infrared data (Mehran and AghaKouchak, 2014).

* *Real-time* usually = there is at most a 1-2 hour delay before observation data is made available raw (i.e. with no gap-filling or other modification).
Near-real-time = there is at most a 1-2 day delay before delivery, allowing some initial data checks to be carried out.
Reanalysis data = data assimilation techniques have been used to fill gaps in the observation data (e.g. missing variables).
Blended = observation data have been combined with either or both of raingauge and reanalysis data to create a more robust and quality-controlled product.

Table 3: Modelling systems details (Dutra et al., 2015; Nikolopoulos et al., 2016). Each model was driven using as close as possible to the same configuration: Global Water Resources Reanalysis 2 (WRR2, Arduini et al. (2017) and http://jules.jchmr.org/content/research-community-configurations). Simulation results are available on the THREDDS data server (https://wci.eartH2Observe.eu/thredds/catalog.html, see Schellekens et al. (2017)).

Model	Institution	Simulations
Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land model (**H-TESSEL**) (Balsamo et al., 2009)	ECMWF	A 10-year spin-up was carried out: an initial run from 1 January 1979 to 1 January 1989, while the land surface state of January 1989 was used to initialize the main simulation.
JULES is the Joint UK Land Environment Simulator model (**JULES**) (Best et al., 2011; Clark et al., 2011)	MetO/CEH	A 10-year spin-up was carried out: an initial run from 1 January 1979 to 1 January 1989, while the land surface state of January 1989 was used to initialize the main simulation.
ORganizing Carbon and Hydrology In Dynamic EcosystEms model (**ORCHIDEE**) (d'Orgeval et al., 2008; Krinner et al., 2005)	CNRS/LSCE	The model was spun up with a simulation from 1 January 1979 to 31 December 1990. This simulation started with an average soil moisture and empty aquifers. After the 12 years of spin-up, river discharges have reached equilibrium.
SURFace EXternalisée model (**SURFEX**) (Decharme et al., 2011; Decharme et al., 2013)	Météo-France	A 20-year spin-up was carried out using the 1979–1988 period twice.
Water – Global Assessment and Prognosis-3 (**WaterGAP3**) (Schneider et al., 2011; Verzano et al., 2012). A grid-based, integrative global fresh water resource assessment tool.	University of Kassel	Storage compartments were initialized by re-running the model with the first year of available meteorological forcing 10 times. WaterGAP includes a water use model (domestic and industrial water use are parameterised as a function of average income per country (GDP/capita), allowing global water use calculations.

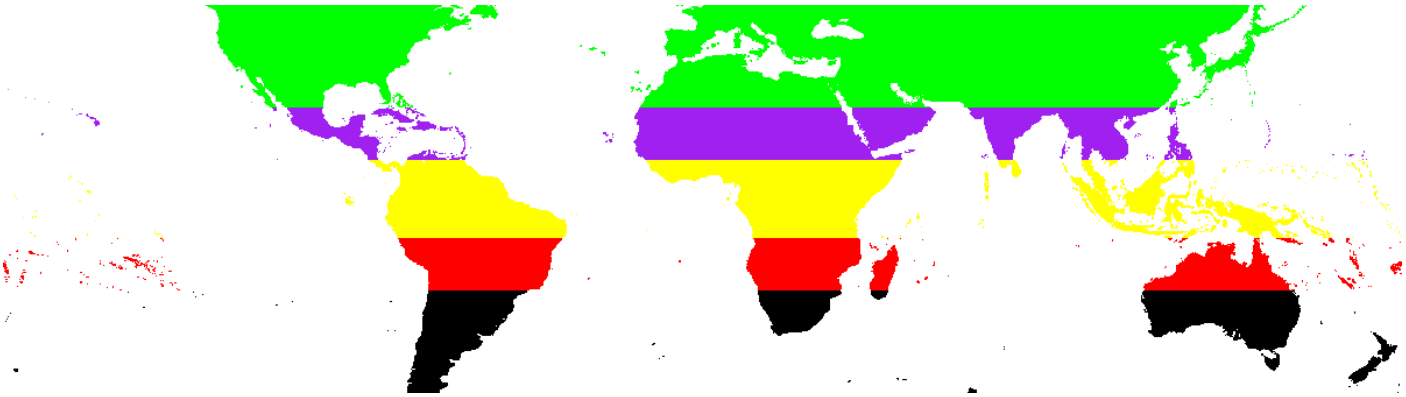


Fig. 1: Latitudinal zones used in this study: black = southern temperate 23.5°S to 50.0°S, red = southern tropical 10.0°S to 23.5°S, yellow = equatorial tropical 10.0°N to 10.0°S, purple = northern tropical 23.5°N to 10.0°N and green = northern temperate 50.0°N to 23.5°S. Analyses are restricted to the area 50.0°N to 50.0°S because of the bounds of data validity in the TRMM and TRMM-RT precipitation data products (Table 2).

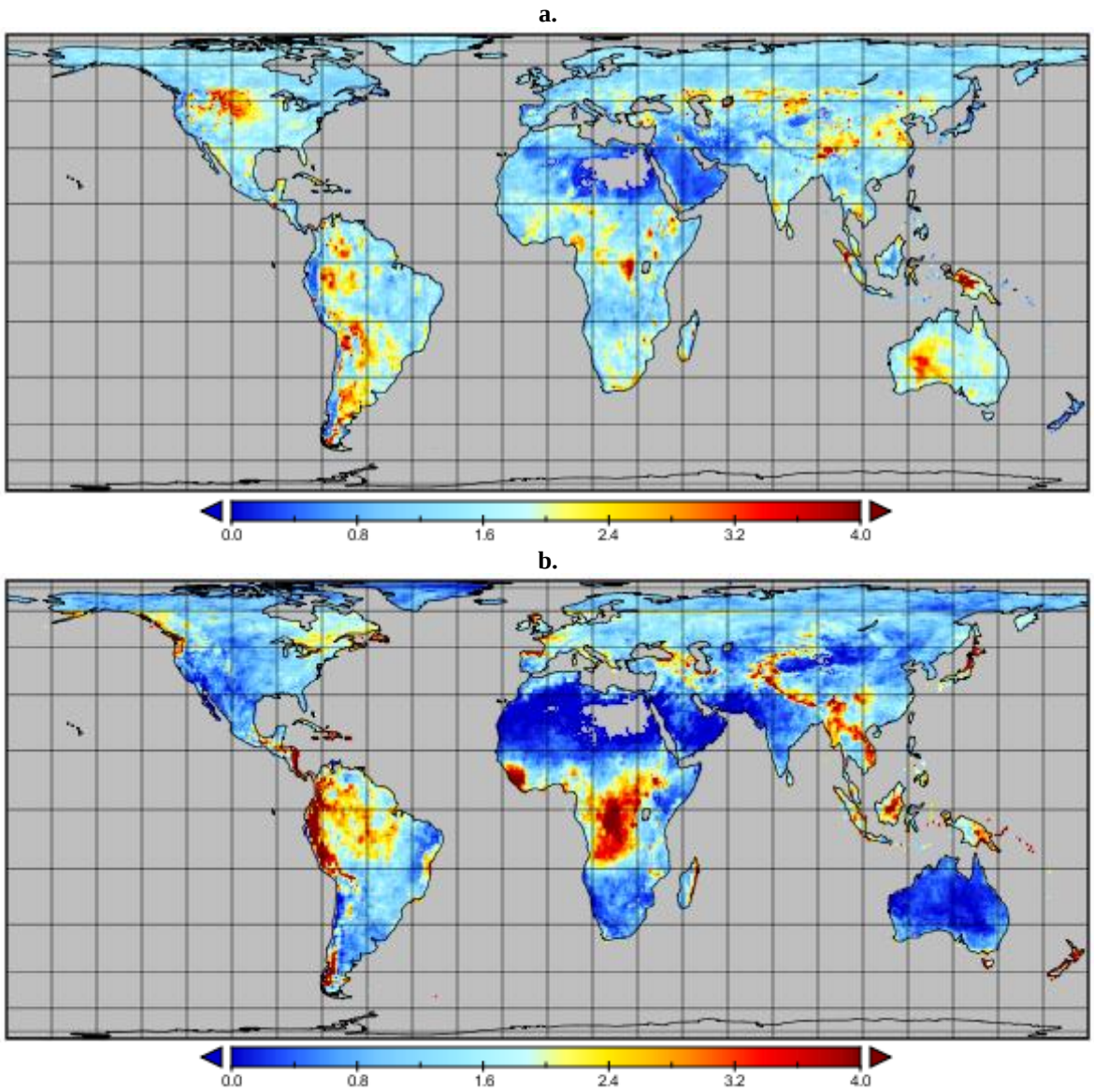


Fig. 2: Uncertainty in the precipitation inputs to the *eartH2Observe* ensemble models: (a) Uncertainty in precipitation extreme highs and (b) Uncertainty in precipitation extreme lows (standard deviation (SD) taken across the precipitation products) in units of (occurrence of extreme events per year). Areas of consistently very low precipitation are masked in grey. Note that only isolated global areas exceeded 4 events/yr, so the scale is restricted to 0-4 events/yr.

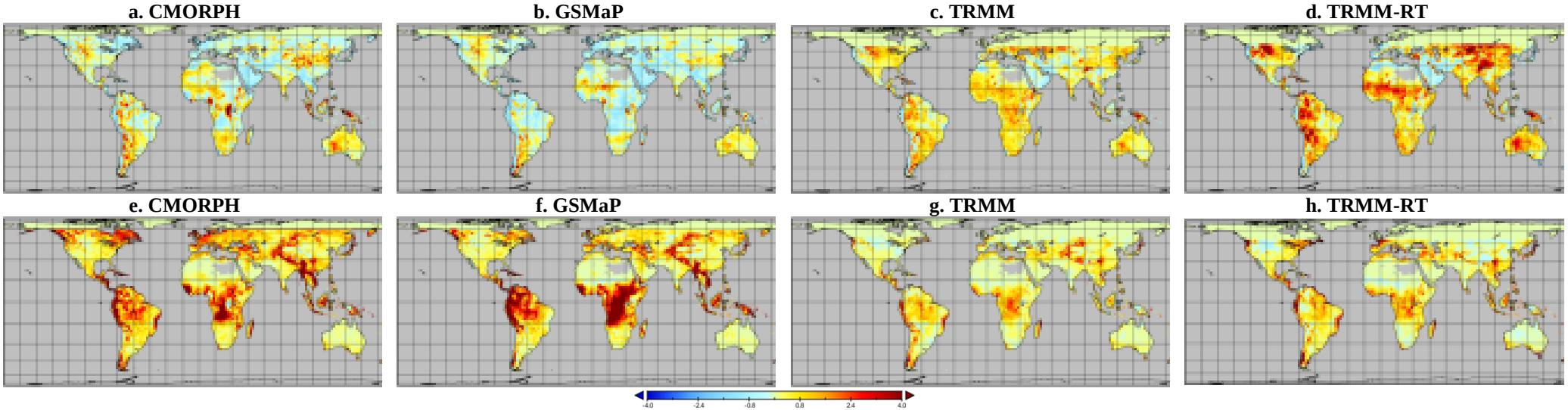
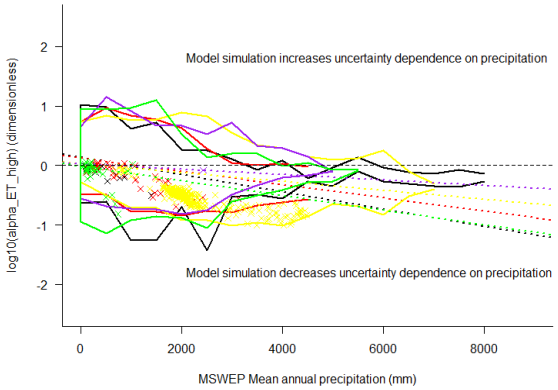
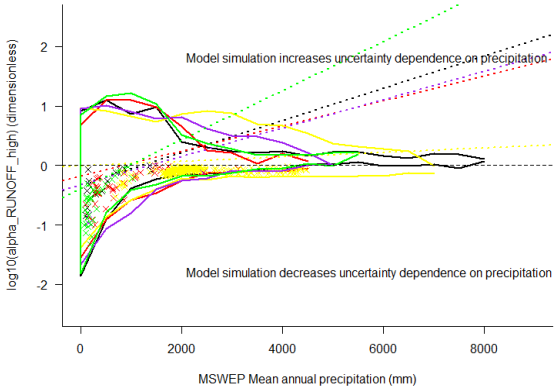


Fig. 3: Increase in extreme precipitation event occurrence in relation to MSWEP. Subtracting extreme high event occurrence rates in the MSWEP precipitation input from the rates in the CMORPH precipitation input gives map (a), and (b) to (d) are the same calculation using GSMaP, TRMM and TRMM-RT instead of CMORPH. (e) to (h) is the same calculation, but for extreme low event occurrence (i.e. the averages of the upper and lower rows are effectively the maps Fig. 2a and Fig. 2b, respectively). The clear lines at 50°N (TRMM, TRMM-RT) and 60°N (CMORPH, GSMaP) show the bounds of data validity for these products (Table 2). Note that only isolated global areas exceeded 4 events/yr, so the scale is restricted to -4 to +4 events/yr.

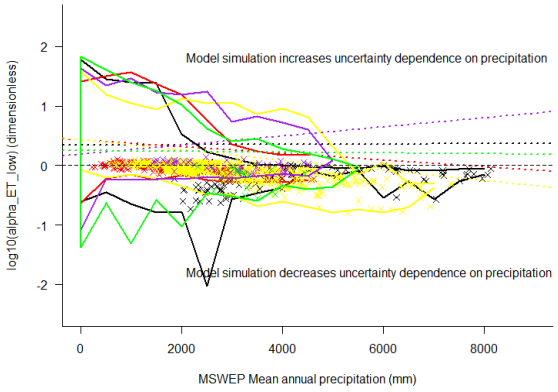
a.


Model simulation increases uncertainty dependence on precipitation
Model simulation decreases uncertainty dependence on precipitation
log10(alpha_ET_high) (dimensionless)
MSWEP Mean annual precipitation (mm)

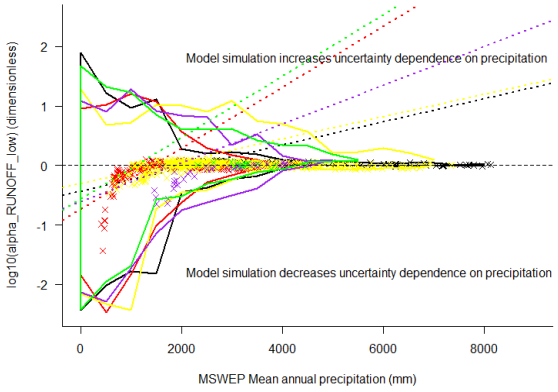
b.


Model simulation increases uncertainty dependence on precipitation
Model simulation decreases uncertainty dependence on precipitation
log10(alpha_RUNOFF_high) (dimensionless)
MSWEP Mean annual precipitation (mm)

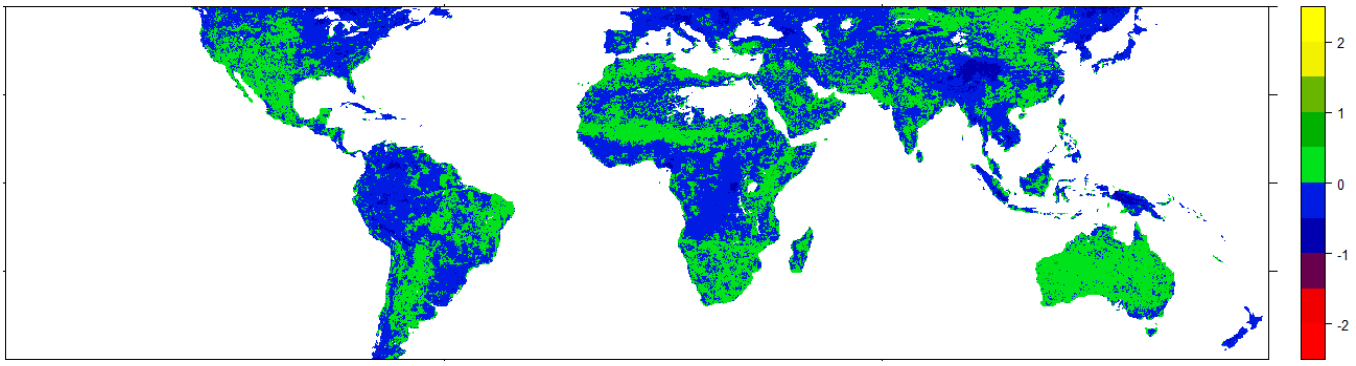
c.


Model simulation increases uncertainty dependence on precipitation
Model simulation decreases uncertainty dependence on precipitation
log10(alpha_ET_low) (dimensionless)
MSWEP Mean annual precipitation (mm)

d.


Model simulation increases uncertainty dependence on precipitation
Model simulation decreases uncertainty dependence on precipitation
log10(alpha_RUNOFF_low) (dimensionless)
MSWEP Mean annual precipitation (mm)

e.



f.

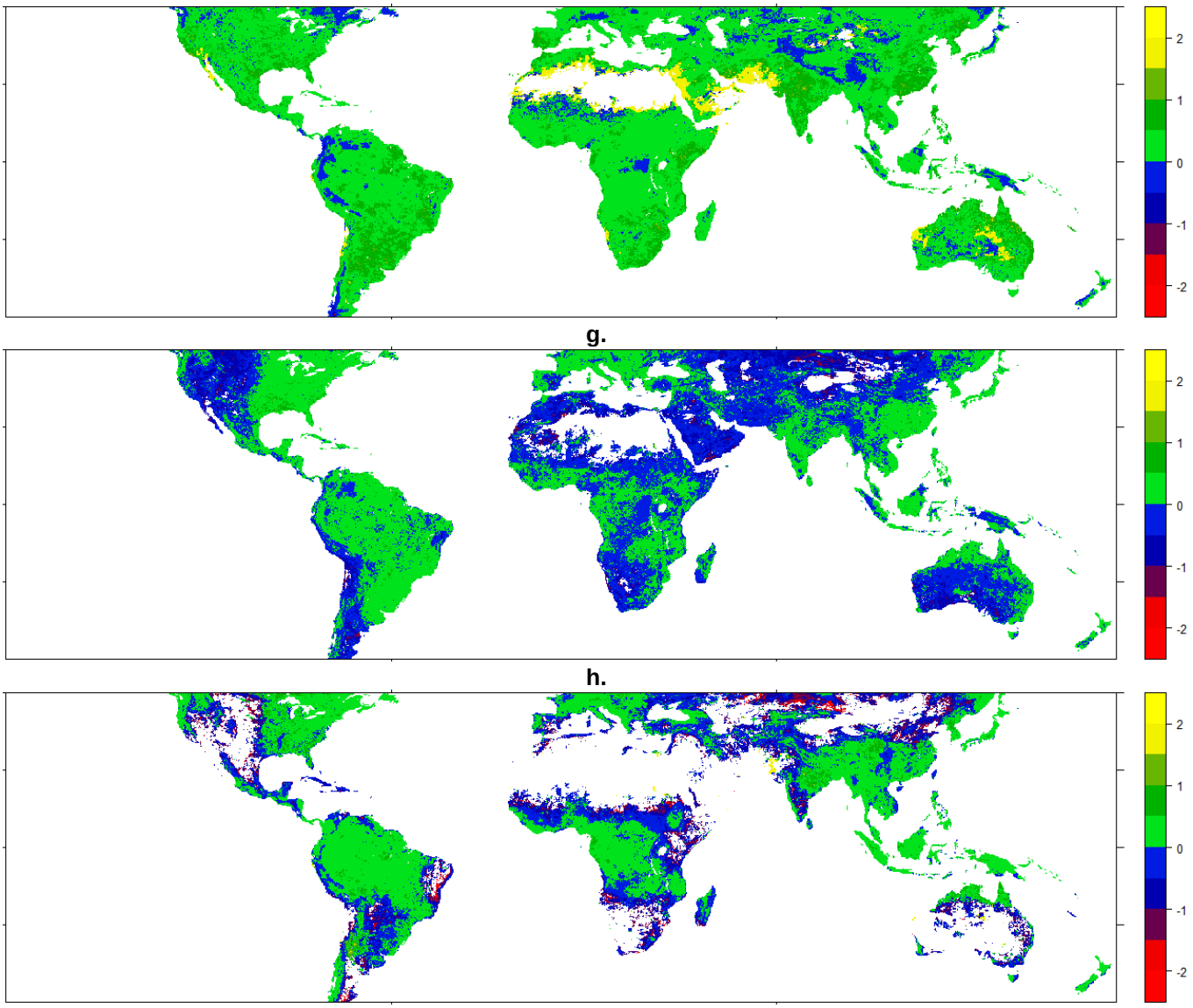
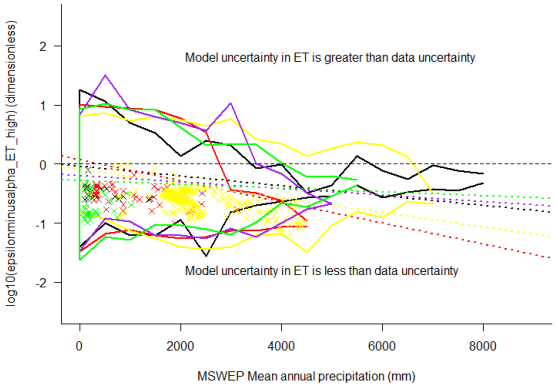
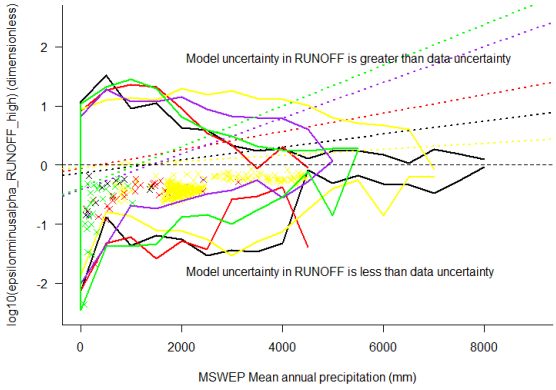


Fig. 4: Values of $\log_{10}(\alpha_{X,j})$, where $\alpha_{X,j}$ is the scaled data uncertainty in variable *X* (eqn 1). ($\log_{10}(\alpha_{X,j})<0$ indicates uncertainty in the predicted variable *X* attributable to the data is less than the variability in the input precipitation forcing data; >0 indicates uncertainty in the predicted variable *X* is greater), where *X* is evapotranspiration (a, c, e, f) or runoff (b, d, g, h) and *j* refers to either high extremes (a, b, e, g) or low extremes (c, d, f, h). Points on the scatter plots are coloured according to latitudinal zones (Fig. 1). Because of the density of overlapping points, only the envelope of points for each latitudinal zone is shown and the points with the highest uncertainty (uncertainty *DIU* ≥ (2/3)*(global maximum of *DIU*)). Linear regression lines for each latitudinal zone indicate the trend as precipitation increases within each zone (all regressions were significant at the 1% level). Maps (e-h) show the corresponding spatial distributions of $\log_{10}(\alpha_{X,j})$ values for each variable, with the colour scales corresponding to the vertical axis on scatter plot (a).

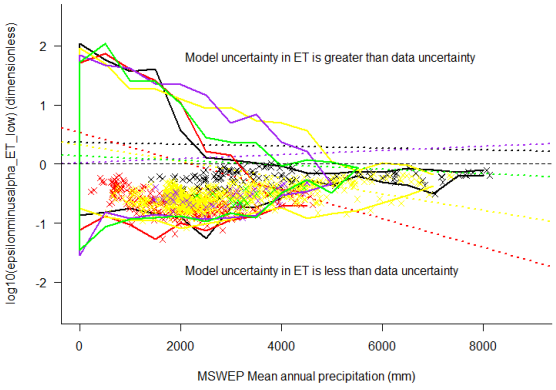
a.


log10(epsilonminusalpha_ET_high) (dimensionless)
Model uncertainty in ET is greater than data uncertainty
Model uncertainty in ET is less than data uncertainty
MSWEP Mean annual precipitation (mm)

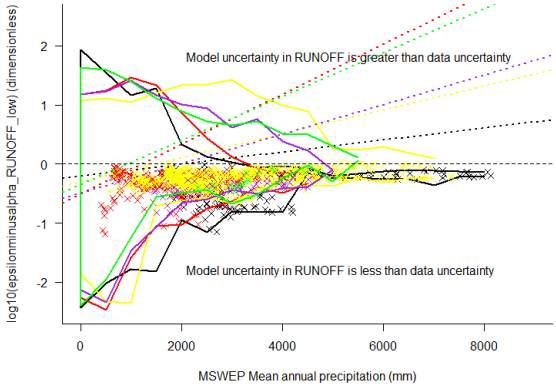
b.


log10(epsilonminusalpha_RUNOFF_high) (dimensionless)
Model uncertainty in RUNOFF is greater than data uncertainty
Model uncertainty in RUNOFF is less than data uncertainty
MSWEP Mean annual precipitation (mm)

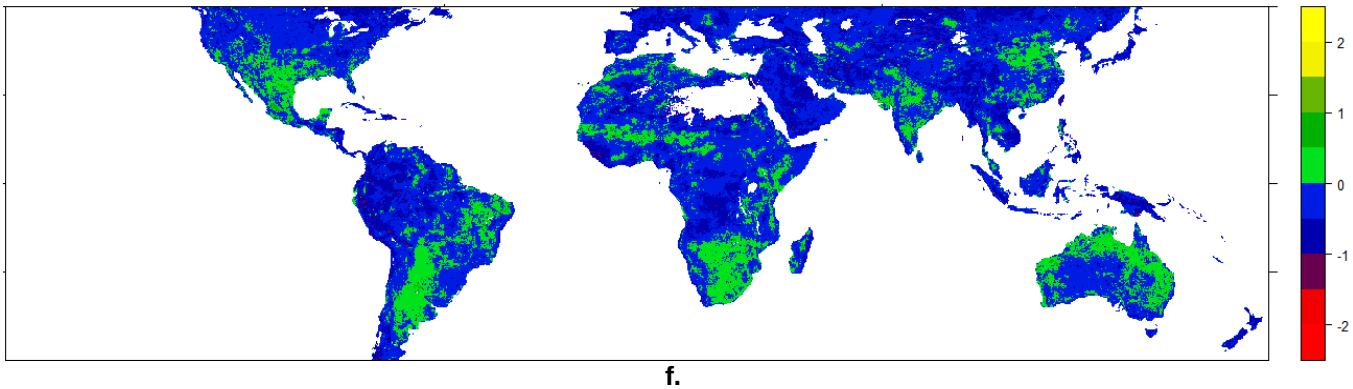
c.


log10(epsilonminusalpha_ET_low) (dimensionless)
Model uncertainty in ET is greater than data uncertainty
Model uncertainty in ET is less than data uncertainty
MSWEP Mean annual precipitation (mm)

d.


log10(epsilonminusalpha_RUNOFF_low) (dimensionless)
Model uncertainty in RUNOFF is greater than data uncertainty
Model uncertainty in RUNOFF is less than data uncertainty
MSWEP Mean annual precipitation (mm)

e.



f.

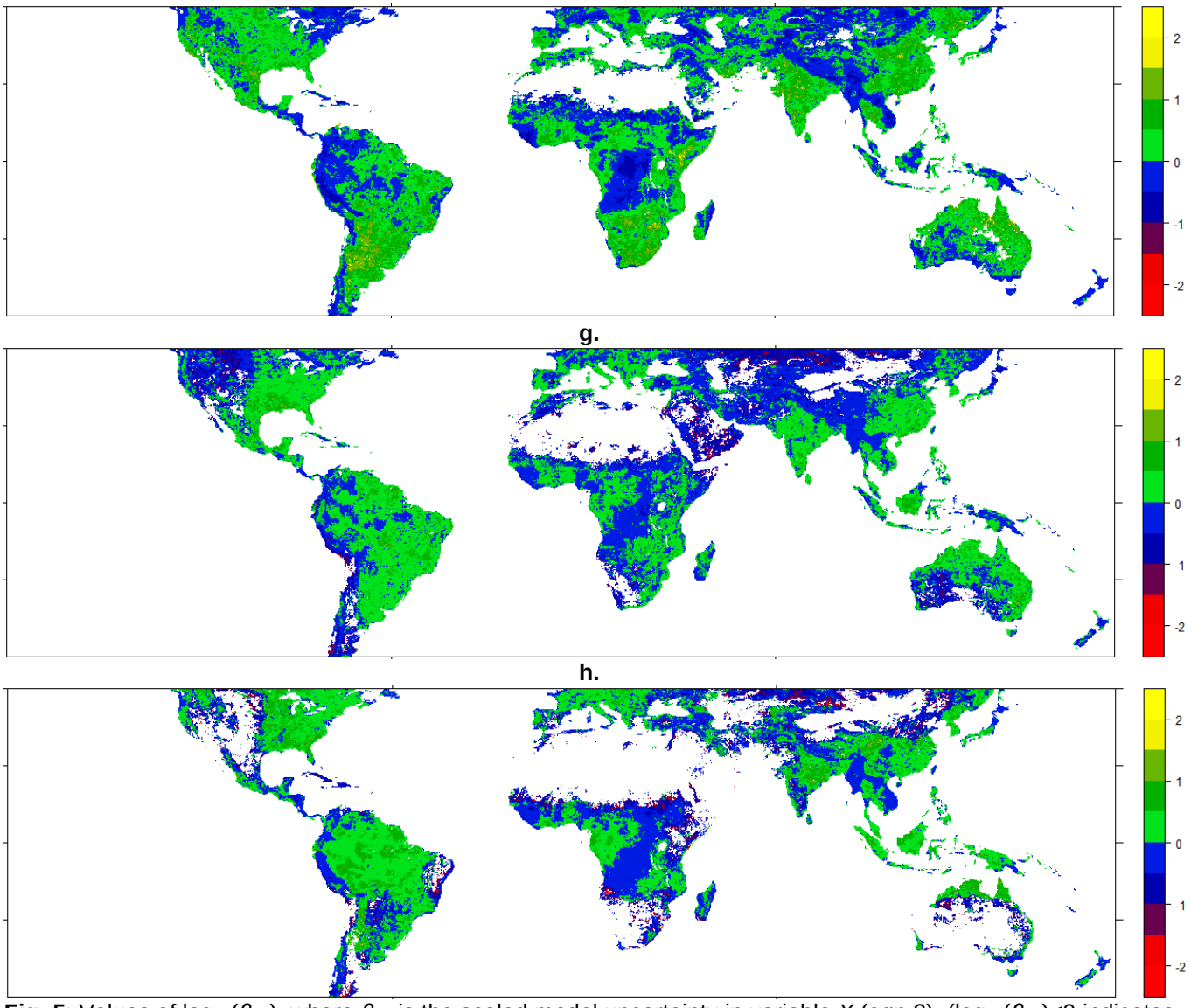
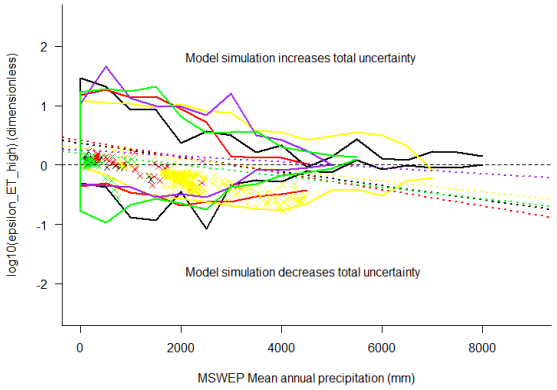
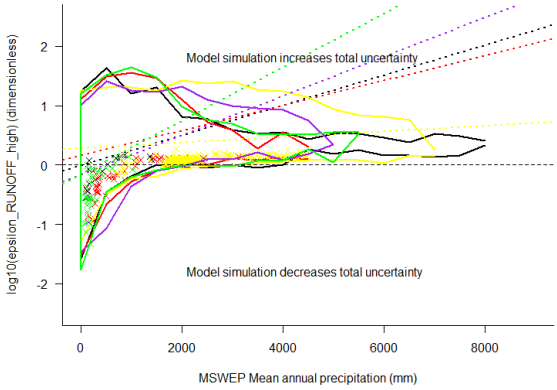


Fig. 5: Values of $\log_{10}(\beta_{X,j})$, where $\beta_{X,j}$ is the scaled model uncertainty in variable *X* (eqn 2). ($\log_{10}(\beta_{X,j})<0$ indicates model uncertainty in the predicted variable *X* is less than the variability in the input precipitation forcing data; >0 indicates model uncertainty in the predicted variable *X* is greater), where *X* is evapotranspiration (a, c, e, f) or runoff (b, d, g, h) and *j* refers to either high extremes (a, b, e, g) or low extremes (c, d, f, h). Points on the scatter plots are coloured according to latitudinal zones (Fig. 1). Because of the density of overlapping points, only the envelope of points for each latitudinal zone is shown and the points with the highest uncertainty (uncertainty *DIU* ≥ (2/3)*(global maximum of *DIU*)). Linear regression lines for each latitudinal zone indicate the trend as precipitation increases within each zone (all regressions were significant at the 1% level). Maps (e-h) show the corresponding spatial distributions of $\log_{10}(\beta_{X,j})$ values for each variable, with the colour scales corresponding to the vertical axis on scatter plot (a).

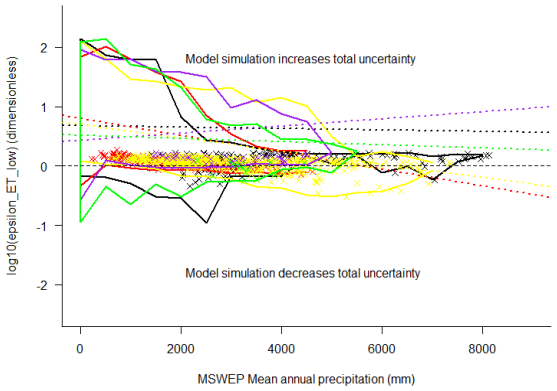
a.


log10(epsilon_ET_high) (dimensionless)
Model simulation increases total uncertainty
Model simulation decreases total uncertainty
MSWEP Mean annual precipitation (mm)

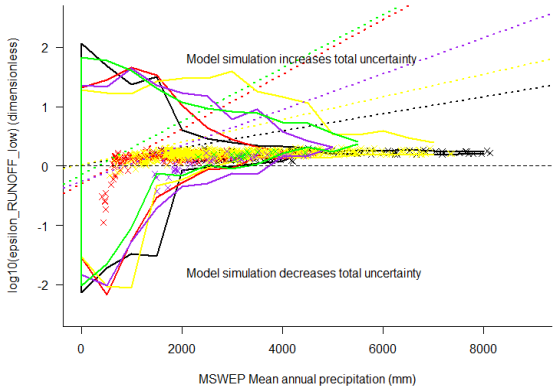
b.


log10(epsilon_RUNOFF_high) (dimensionless)
Model simulation increases total uncertainty
Model simulation decreases total uncertainty
MSWEP Mean annual precipitation (mm)

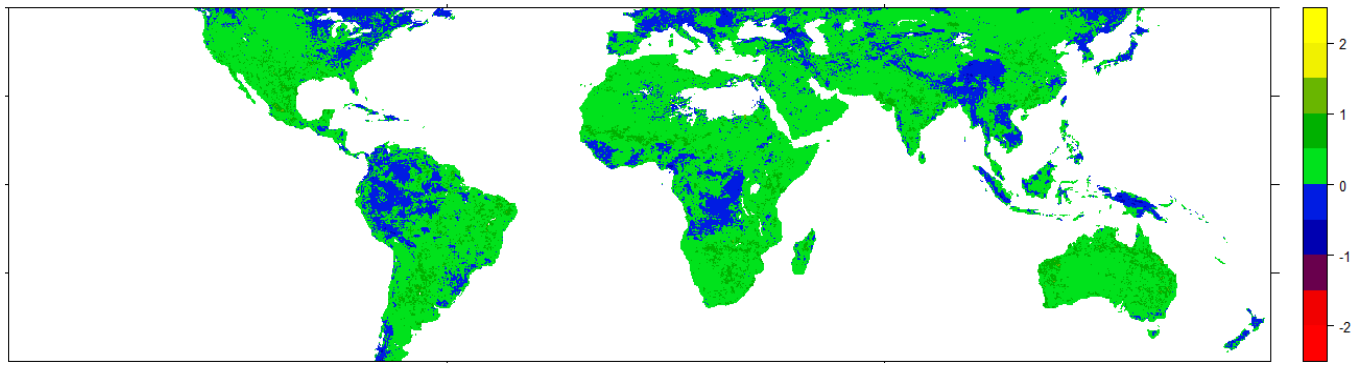
c.


log10(epsilon_ET_low) (dimensionless)
Model simulation increases total uncertainty
Model simulation decreases total uncertainty
MSWEP Mean annual precipitation (mm)

d.


log10(epsilon_RUNOFF_low) (dimensionless)
Model simulation increases total uncertainty
Model simulation decreases total uncertainty
MSWEP Mean annual precipitation (mm)

e.



f.

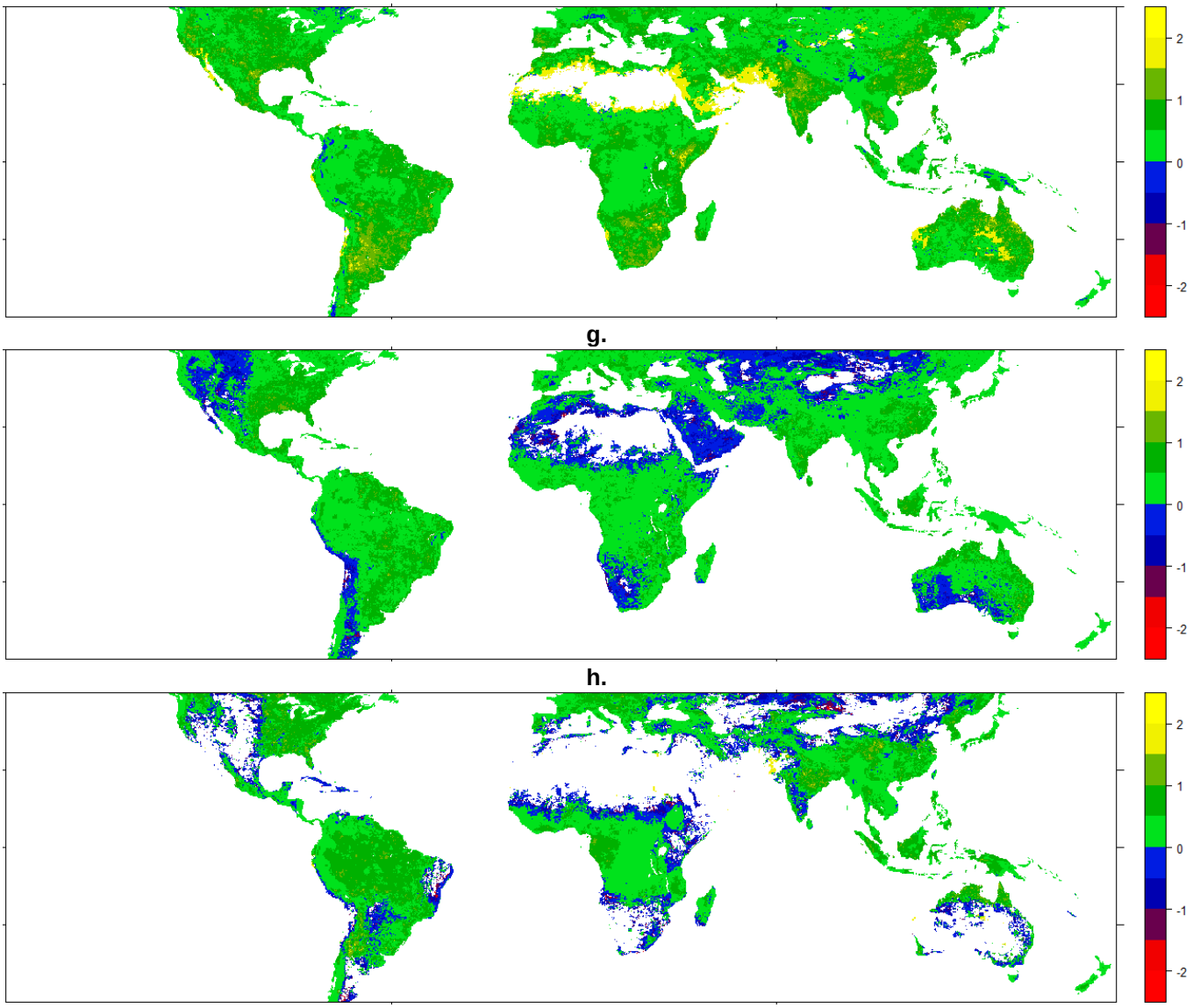


Fig. 6: Values of $\log_{10}(\varepsilon_{X,j})$, where $\varepsilon_{X,j}$ is the total uncertainty in variable *X* (eqn 3), where *X* is evapotranspiration (a, c, e, f) or runoff (b, d, g, h) and *j* refers to either high extremes (a, b, e, g) or low extremes (c, d, f, h). Points on the scatter plots are coloured according to latitudinal zones (Fig. 1). Because of the density of overlapping points, only the envelope of points for each latitudinal zone is shown and the points with the highest uncertainty (uncertainty *DIU* ≥ (2/3)*(global maximum of *DIU*)). Linear regression lines for each latitudinal zone indicate the trend as precipitation increases within each zone (all regressions were significant at the 1% level). Maps (e-h) show the corresponding spatial distributions of $\log_{10}(\varepsilon_{X,j})$ values for each variable, with the colour scales corresponding to the vertical axis on scatter plot (a).

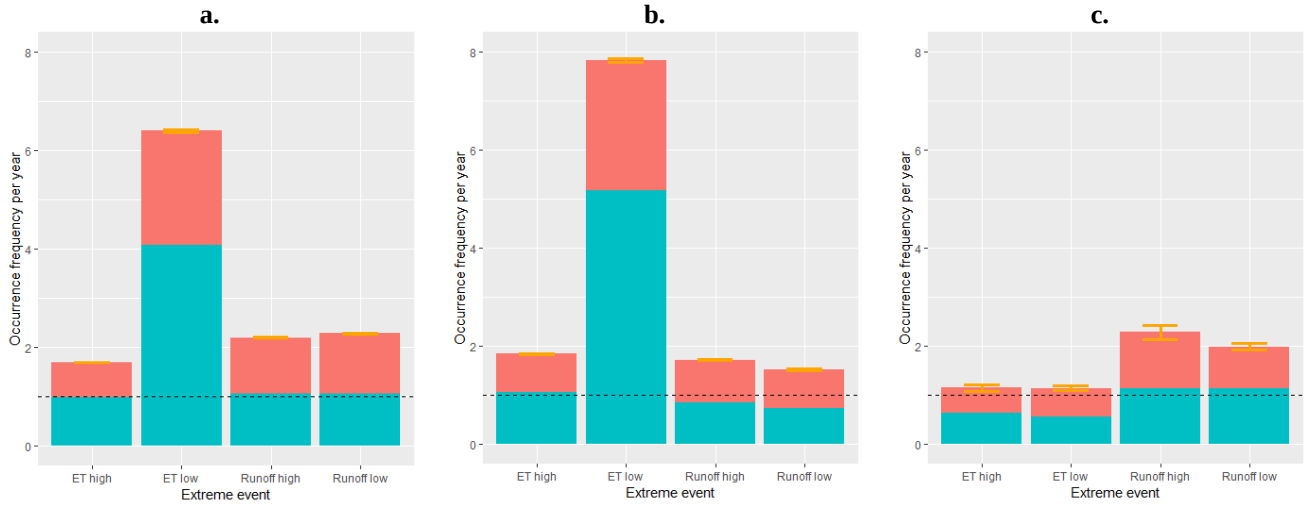


Fig. 7: Global mean values (averaged over 50°S to 50°N) from scatter plots in Fig. 4, Fig. 5 and Fig. 6. Plots show (a) all values, (b) values from dry environments with mean annual precipitation <1000 mm/yr only and (c) values from wet environments ≥6000 mm/yr only. Bar heights are ε values (scaled total uncertainty), with ■ showing α values (scaled data uncertainty) and ■ β (scaled model uncertainty); error bars show SE.

Acknowledgements

We gratefully acknowledge funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 603608, Global Earth Observation for integrated water resource assessment: *eartH2Observe*.

References

Ali, H. and Mishra, V.: Contrasting response of rainfall extremes to increase in surface air and dewpoint temperatures at urban locations in India, Scientific reports, 7, https://doi.org/10.1038/S41598-017-01306-1, 2017.

Arduini, G., Boussetta, S., Dutra, E., and Martínez de la Torre, A.: Report on the Ensemble Water Resources Reanalysis, 2017.

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A. K., and Scipal, K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J Hydrometeorol, 10, 623-643, https://doi.org/10.1175/2008JHM1068.1, 2009.

Bates, B., Kundzewicz, Z. W., Wu, S., and Palutikof, J.: Climate Change and Water. In: IPCC Technical Paper 6, Geneva, Switzerland, 2008.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25 degrees global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrol Earth Syst Sc, 21, 589-615, https://doi.org/10.5194/hess-21-589-2017, 2017a.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol Earth Syst Sc, 21, 6201-6217, https://doi.org/10.5194/hess-21-6201-2017, 2017b.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677-699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Bhuiyan, M. A. E., Nikolopoulos, E. I., Anagnostou, E. N., Albergel, C., Dutra, E., Fink, G., Martínez de la Torre, A., Munier, S., and Polcher, J.: Assessment of Precipitation Error Propagation in Multi-model Global Water Resources Reanalysis, Hydrology and Earth System Sciences Discussions, doi: 10.5194/hess-2018-434, **2018a**. https://doi.org/10.5194/hess-2018-434, 2018a.

Bhuiyan, M. A. E., Nikolopoulos, E. I., Anagnostou, E. N., Quintana-Segui, P., and Barella-Ortiz, A.: A nonparametric statistical technique for combining global precipitation datasets: development and hydrological evaluation over the Iberian Peninsula, Hydrol Earth Syst Sc, 22, 1371-1389, https://doi.org/10.5194/hess-22-1371-2018, 2018b.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour Res, 51, 4923-4947, https://doi.org/10.1002/2015WR017173, 2015.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701-722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

d'Orgeval, T., Polcher, J., and de Rosnay, P.: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, Hydrol Earth Syst Sc, 12, 1387-1401, https://doi.org/DOI 10.5194/hess-12-1387-2008, 2008.

Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., Heathwaite, L., Holden, J., Holman, I. P., Lane, S. N., O'Connell, E., Penning-Rowsell, E., Reynard, N., Sear, D., Thorne, C., and Wilby, R.: A restatement of the natural science evidence concerning catchment-based 'natural' flood management in the UK, P Roy Soc a-Math Phy, 473, https://doi.org/10.1098/Rspa.2016.0706, 2017.

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions, J Geophys Res-Atmos, 116, https://doi.org/10.1029/2011jd016002, 2011.

Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J Geophys Res-Atmos, 118, 7819-7834, https://doi.org/10.1002/jgrd.50631, 2013.

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C. H., Delrieu, G., Hong, Y., Hsu, Y. C., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E. I., Sahlu, D., Salerno, F., Rodriguez-Sanchez, J. P., Vergara, H. J., and Yilmaz, K. K.: Multiregional Satellite Precipitation Products Evaluation over Complex Terrain, J Hydrometeorol, 17, 1817-1836, 2016.

Dutra, E., Balsamo, G., Calvet, J., Minvielle, M., Eisner, S., Fink, G., Peβenteiner, S., Orth, R., Burke, S., van Dijk, A., Polcher, J., Beck, H., and Martínez-de la Torre, A.: Report on the current state-of-the-art Water Resources Reanalysis, 2015.

He, J., Deser, C., and Soden, B. J.: Atmospheric and Oceanic Origins of Tropical Precipitation Variability, J Climate, 30, 3197-3217, https://doi.org/10.1175/Jcli-D-16-0714.1, 2017.

Hirpa, F. A., Salamon, P., Alfieri, L., Thielen-del Pozo, J., Zsoter, E., and Pappenberger, F.: The Effect of Reference Climatology on Global Flood Forecasting, J Hydrometeorol, 17, 1131-1145, https://doi.org/10.1175/Jhm-D-15-0044.1, 2016.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G. J., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J Hydrometeorol, 8, 38-55, https://doi.org/10.1175/JHM560.1, 2007.

Huntingford, C., Zelazowski, P., Galbraith, D., Mercado, L. M., Sitch, S., Fisher, R., Lomas, M., Walker, A. P., Jones, C. D., Booth, B. B. B., Malhi, Y., Hemming, D., Kay, G., Good, P., Lewis, S. L., Phillips, O. L., Atkin, O. K., Lloyd, J., Gloor, E., Zaragoza-Castells, J., Meir, P., Betts, R., Harris, P. P., Nobre, C., Marengo, J., and Cox, P. M.: Simulated resilience of tropical rainforests to CO2-induced climate change, Nat Geosci, 6, 268-273, https://doi.org/10.1038/NGEO1741, 2013.

IPCC: Climate Change 2014: The Physical Science Basis, Intergovernmental Panel on Climate Change (IPCC), U.K., 2014.

IPCC: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, Cambridge, UK, 2012.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P. P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, J Hydrometeorol, 5, 487-503, https://doi.org/Doi 10.1175/1525-7541(2004)005<0487:Camtpg>2.0.Co;2, 2004.

Kimani, M. W., Hoedjes, J. C. B., and Su, Z. B.: An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa, Remote Sens-Basel, 9, https://doi.org/10.3390/Rs9050430, 2017.

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, https://doi.org/10.1029/2003gb002199, 2005.

Kundzewicz, Z. W. and Stakhiv, E. Z.: Are climate models "ready for prime time" in water resources management applications, or is more research needed?, Hydrolog Sci J, 55, 1085-1089, https://doi.org/10.1080/02626667.2010.513211, 2010.
Levizzani, V., Kidd, C., Aonashi, K., Bennartz, R., Ferraro, R. R., Huffman, G. J., Roca, R., Turk, F. J., and Wang, N. Y.: The activities of the International Precipitation Working Group, Quarterly Journal of the Royal Meteorological Society, 144, 3-15, https://doi.org/10.1002/qj.3214, 2018.
Li, H. B. and Wu, J. G.: Uncertainty analysis in ecological studies: An overview. In: Scaling and Uncertainty Analysis in Ecology: Methods and Applications, Springer, Netherlands, 2006.
Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, Water Resour Res, 50, 1131-1151, https://doi.org/10.1002/2013wr014581, 2014.
Lopez, O., Houborg, R., and McCabe, M. F.: Evaluating the hydrological consistency of evaporation products using satellite-based gravity and rainfall data, Hydrol Earth Syst Sc, 21, 323-343, https://doi.org/10.5194/hess-21-323-2017, 2017.
Luo, Z. J., Anderson, R. C., Rossow, W. B., and Takahashi, H.: Tropical cloud and precipitation regimes as seen from near-simultaneous TRMM, CloudSat, and CALIPSO observations and comparison with ISCCP, Journal of Geophysical Research: Atmospheres, 122, 5988-6003, 2017.
Marshall, T. J., Holmes, J. W., and Rose, C. W.: Soil Physics, Cambridge University Press, Cambridge, U.K., 1996.
Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903-1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.
Marthews, T. R., Malhi, Y., Girardin, C. A., Silva Espejo, J. E., Aragao, L. E., Metcalfe, D. B., Rapp, J. M., Mercado, L. M., Fisher, R. A., Galbraith, D. R., Fisher, J. B., Salinas-Revilla, N., Friend, A. D., Restrepo-Coupe, N., and Williams, R. J.: Simulating forest productivity along a neotropical elevational transect: temperature variation and carbon use efficiency, Global change biology, 18, 2882-2898, https://doi.org/10.1111/j.1365-2486.2012.02728.x, 2012.
McGregor, G. R. and Nieuwolt, S.: Tropical Climatology, Wiley, Chichester, UK, 1998.
Mehran, A. and AghaKouchak, A.: Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations, Hydrological Processes, 28, 2262-2270, https://doi.org/10.1002/hyp.9779, 2014.
Munier, S., Minvielle, M., Decharme, B., Calvet, J., Blyth, E., Veldkamp, T. I. E., and Nikolopolous, T.: Report on uncertainty characterization of the WP5 WRR tier 2 products, 2018.
Nikolopoulos, E., Anagnostou, M., Albergel, C., Dutra, E., Fink, G., Martínez-de la Torre, A., Munier, S., Polcher, J., and Quintana-Segui, P.: Report on precipitation error modeling and ensemble error propagation using LSM and GHM models from tier 1 reanalysis, 2016.
Oberkampf, W. L. and Roy, C. J.: Verification and Validation in Scientific Computing, Cambridge University Press, 2010.
Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tolle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev Geophys, 53, 323-361, https://doi.org/10.1002/2014RG000475, 2015.
Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, P Natl Acad Sci USA, 111, 3262-3267, https://doi.org/10.1073/pnas.1222473110, 2014.
Riley, K., Thompson, M., Webley, P., and Hyde, K. D.: Uncertainty in Natural Hazards, Modeling and Decision Support. In: Natural Hazard Uncertainty Assessment: Modeling and-Decision Support, Riley, K., Webley, P., and Thompson, M. (Eds.), Wiley, 2017.
Schellekens, J., Dutra, E., Martinez-de la Torre, A., Balsamo, G., van Dijk, A., Weiland, F. S., Minvielle, M., Calvet, J. C., Decharme, B., Eisner, S., Fink, G., Florke, M., Pessenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth Syst Sci Data, 9, 389-413, https://doi.org/10.5194/essd-9-389-2017, 2017.
Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colon-Gonzalez, F. J., Gosling, S. N., Kim, H., Liu, X. C., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q. H., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, P Natl Acad Sci USA, 111, 3245-3250, https://doi.org/10.1073/pnas.1222460110, 2014.
Schneider, C., Florke, M., Eisner, S., and Voss, F.: Large scale modelling of bankfull flow: An example for Europe, J Hydrol, 408, 235-245, https://doi.org/10.1016/j.jhydrol.2011.08.004, 2011.
Taylor, C. M., de Jeu, R. A., Guichard, F., Harris, P. P., and Dorigo, W. A.: Afternoon rain more likely over drier soils, Nature, 489, 423-426, https://doi.org/10.1038/nature11377, 2012.
Tian, Y. D., Peters-Lidard, C. D., Adler, R. F., Kubota, T., and Ushio, T.: Evaluation of GSMaP Precipitation Estimates over the Contiguous United States, J Hydrometeorol, 11, 566-574, https://doi.org/10.1175/2009JHM1190.1, 2010.
Trenberth, K. E., Fasullo, J. T., and Shepherd, T. G.: Attribution of climate extreme events, Nature Climate Change, 5, 725-730, https://doi.org/10.1038/nclimate2657, 2015.

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., Decker, M., and Haughton, N.: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts, Environmental Research Letters, 11, 104012, https://doi.org/10.1088/1748-9326/11/10/104012, 2016.

Van Loon, A. F.: Hydrological drought explained, Wires Water, 2, 359-392, https://doi.org/10.1002/wat2.1085, 2015.

Veldkamp, T. I. E. and Ward, P.: Report on global scale assessment of physical and social water scarcity, 2015.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environmental Research Letters, 13, 2018.

Verzano, K., Barlund, I., Florke, M., Lehner, B., Kynast, E., Voss, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, J Hydrol, 424, 238-251, https://doi.org/10.1016/j.jhydrol.2012.01.005, 2012.

Wartenburger, R., Seneviratne, S. I., Hirschi, M., Chang, J. F., Ciais, P., Deryng, D., Elliott, J., Folberth, C., Gosling, S. N., Gudmundsson, L., Henrot, A. J., Hickler, T., Ito, A., Khabarov, N., Kim, H., Leng, G. Y., Liu, J. G., Liu, X. C., Masaki, Y., Morfopoulos, C., Muller, C., Schmied, H. M., Nishina, K., Orth, R., Pokhrel, Y., Pugh, T. A. M., Satoh, Y., Schaphoff, S., Schmid, E., Sheffield, J., Stacke, T., Steinkamp, J., Tang, Q. H., Thiery, W., Wada, Y., Wang, X. H., Weedon, G. P., Yang, H., and Zhou, T.: Evapotranspiration simulations in ISIMIP2a-Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets, Environmental Research Letters, 13, 2018.

Yi, C., Pendall, E., and Ciais, P.: Focus on extreme events and the carbon cycle, Environmental Research Letters, 10, 070201, https://doi.org/10.1088/1748-9326/10/7/070201, 2015.

Yin, H., Donat, M. G., Alexander, L. V., and Sun, Y.: Multi-dataset comparison of gridded observed temperature and precipitation extremes over China, Int J Climatol, 35, 2809-2827, https://doi.org/10.1002/joc.4174, 2015.