Ground truthing global-scale model estimates of groundwater recharge across Africa

Charles West [a,*], Robert Reinecke [b], Rafael Rosolem [a,c], Alan M. MacDonald [d], Mark O. Cuthbert [e,f], Thorsten Wagener [a,b]

[a] Civil Engineering, University of Bristol, Bristol, United Kingdom
[b] Institute for Environmental Science and Geography, University of Potsdam, 14476 Potsdam, Germany
[c] Cabot Institute for the Environment, University of Bristol, Bristol, United Kingdom
[d] British Geological Survey, Lyell Centre, Edinburgh EH14 4AP, United Kingdom
[e] School of Earth and Environmental Sciences, Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom
[f] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, Australia

HIGHLIGHTS

- Model estimates of annual recharge and recharge ratio disagree across Africa
- Climate controls mostly explain the spatial variability in model recharge estimates
- Model similarity to ground-based estimates varies inconsistently throughout Africa

GRAPHICAL ABSTRACT
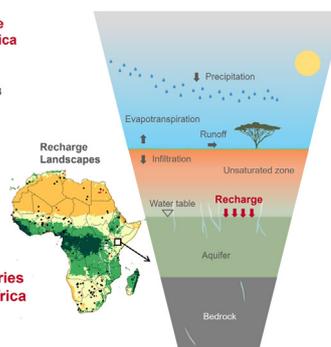
ARTICLE INFO

ABSTRACT

Groundwater is an essential resource for natural and human systems throughout the world and the rates at which aquifers are recharged constrain sustainable levels of consumption. However, recharge estimates from global-scale models regularly disagree with each other and are rarely compared to ground-based estimates. We compare long-term mean annual recharge and recharge ratio (annual recharge/annual precipitation) estimates from eight global models with over 100 ground-based estimates in Africa. We find model estimates of annual recharge and recharge ratio disagree significantly across most of Africa. Furthermore, similarity to ground-based estimates between models also varies considerably and inconsistently throughout the different landscapes of Africa. Models typically showed both positive and negative biases in most landscapes, which made it challenging to pinpoint how recharge prediction by global-scale models can be improved. However, global-scale models which reflected stronger climatic controls on their recharge estimates compared more favourably to ground-based estimates. Given this significant uncertainty in recharge estimates from current global-scale models, we stress that groundwater recharge prediction across Africa, for both research investigations and operational management, should not rely upon estimates from a single model but instead consider the distribution of estimates from different models. Our work will be of particular interest to decision makers and researchers who consider using such recharge outputs to make groundwater governance decisions or investigate groundwater security especially under the potential impact of climate change.

* Corresponding author.
  E-mail address: charliewest1992@hotmail.co.uk (C. West).

1. Introduction

Global groundwater resources dwarf all alternative sources of freshwater (Gleeson et al., 2016) and are used by billions of people worldwide for drinking water, for securing global food production (Dalin et al., 2017), and for industrial water supply. As a key component in determining sustainable groundwater use (Gleeson et al., 2020), understanding how groundwater recharge varies in space and time is essential for ensuring groundwater security (Aeschbach-Hertig and Gleeson, 2012; MacDonald et al., 2021). However, we cannot directly measure this critical variable across larger domains (or arguably even at small scales); thus, continental and global-scale investigations into groundwater security and variability rely on groundwater recharge estimates from commonly used global-scale models. Recharge estimates from these models have helped to investigate global groundwater depletion (Döll et al., 2014; Wada et al., 2010), its key drivers (Dalin et al., 2017; Döll, 2009), and alternative constraints on sustainable usage (Gleeson et al., 2012; de Graaf et al., 2019). Similarly, de Graaf et al. (2015) and Reinecke et al. (2019) used recharge outputs from a global-scale model to drive global gradient-based groundwater models. Cuthbert et al. (2019) used recharge estimates from a single global-scale model to explore global patterns in climate-groundwater interactions, though there they also approximated the range of likely uncertainty in recharge by comparison to a second global recharge model data set.

However, recharge outputs from these models are poorly- or unconstrained as such models are typically either uncalibrated or calibrated against streamflow records (de Graaf et al., 2015; Döll and Fiedler, 2008). Calibration of global-scale models to groundwater heads, as done by de Graaf et al. (2017), suffers from commensurability issues due to scale differences between observations and modelled variables (Gleeson et al., 2021; Reinecke et al., 2020). Furthermore, any model calibration is typically biased towards data-rich regions of the world such as the USA and Europe (Döll and Fiedler, 2008). Therefore, in data-sparse regions such as Africa, recharge estimates are more dependent upon choice of model structures and uncalibrated parameterisations directly derived from global datasets of relevant soils and other properties. Thus, such models typically reflect a limited representation of those processes not easily captured in globally available data (Wagener et al., 2021), e.g., subsurface heterogeneity produced by karst (Hartmann et al., 2017), or transmission losses in drylands (Quichimbo et al., 2021). Uncertainties in these model structures and parameterisations are then propagated into subsequent analyses, which rarely quantify the influence of potentially divergent recharge estimates between models (Reinecke et al., 2021). Instead, authors usually select recharge outputs from a single model (Wada et al., 2010; de Graaf et al., 2015; Reinecke et al., 2019). If recharge outputs from several models disagree considerably, this likely suggests that our current understanding of groundwater behaviour and security attained by subsequent analyses is uncertain and possibly not robust.

Syntheses of ground-based estimates compiled from the literature could help us understand whether global model estimates appear reasonable or at least plausible. Recharge rates in these ground-based studies are estimated by an array of different methods with different strengths and weaknesses, including environmental tracers, chloride mass balance, water table fluctuations, soil moisture balance models, water balance models, and groundwater models (Scanlon et al., 2002). Mohan et al. (2018), Moeck et al. (2020) and MacDonald et al. (2021) have recently used literature compilation datasets to understand how and why recharge rates vary across continental to global scales. However, these ground-based estimates have rarely been used to assess whether recharge outputs from global-scale models are reasonable.

Döll and Fiedler (2008) provide an example of how such estimates can be used to evaluate global-scale models, comparing annual recharge rates estimated by the global hydrologic model WaterGAP to those from 25 chloride profiles in arid and semi-arid regions. The authors found that WaterGAP overestimated recharge in dry settings and subsequently adjusted the model in these environments to only allow recharge on days when rainfall exceeded 10 mm. Jasechko et al. (2014) used ground-based

estimates to evaluate another global model, PCR-GLOBWB, assessing whether seasonal biases in recharge ratios (recharge/precipitation) agreed with those identified using groundwater isotopes. However, neither study evaluated the models according to their ability convert reasonable proportions of rainfall into recharge. Doing so could highlight whether discrepancies are associated with the partitioning of rainfall to recharge (i.e., model structure) or perhaps differences in precipitation forcing.

In this study, we compare, for the first time, groundwater recharge estimates of eight global hydrological models (Reinecke et al., 2021) with a recently compiled dataset of over 100 ground-based estimates of groundwater recharge across the African continent (MacDonald et al., 2021). By comparing multiple models to ground-based estimates, we can examine where different modelling frameworks seem plausible. We focus on Africa as it is a data sparse region where the use of large-scale hydrological models is particularly important and because groundwater is essential for water security in many parts of the continent (MacDonald et al., 2012; Calow et al., 2010). Three research questions guide our investigations into how model structure uncertainty influences recharge estimates from global-scale models in data-sparse regions.

- Where do model estimates of average annual recharge and recharge ratio agree or disagree?
- Which environmental controls describe the recharge patterns produced by the different models?
- How do model estimates compare to ground-based estimates compiled from the literature?

2. Data and methods

In this section we discuss the key datasets and methods used in our analysis. This includes ground-based estimates of groundwater recharge from the literature, groundwater recharge outputs from global-scale models and the use of Random Forests to investigate the environmental controls on recharge outputs from global-scale models.

2.1. Global models

We compare historical (1979–2005) recharge estimates from eight global-scale models within the Inter-sectoral Impact Model Intercomparison Project (ISIMIP), simulation round 2b (Reinecke et al., 2021). ISIMIP is a model intercomparison framework that enables the comparison of climate impact projections in different sectors (Frieler et al., 2017). We use the recharge estimates available through this project, all of which are provided at the same 0.5° × 0.5° uniform grid (approx. 50 × 50 km at the equator), and do not run the model simulations ourselves. Six of the eight models incorporated time-varying historical land use and water abstractions in their simulations. Two took a different approach; for CLM 4.5, abstractions and land use are fixed to 2005, and JULES-W1 does not model any abstractions. Telteu et al. (2021) provide a complete description of all the models included within ISIMIP and Reinecke et al. (2021) discuss how each model calculates groundwater recharge. Below we summarize the relevant model features reported in each paper (Table 1).

All model simulations used in this study are driven by the HadGEM2-ES Global Circulation Model, developed by the UK Met Office Hadley Centre (Collins et al., 2008). The Global Circulation Model was bias-adjusted by Frieler et al. (2017) using a trend preserving algorithm and the EWEMBI data as a baseline climate condition (Lange, 2018).

2.2. Ground-based groundwater recharge estimates

Ground-based recharge estimates in Africa were initially compiled from the literature by (MacDonald et al., 2021) and are shown here in Fig. 4. MacDonald et al. (2021) undertook a thorough quality assurance when compiling the dataset which includes comprehensive meta-information such as uncertainty ranges on the recharge estimates. For these reasons,

Table 1

List of ISIMIP 2b global-scale models included in our analysis with details about some structural properties, recharge definition and calibration procedure. Model type abbreviations stand for Global Hydrological Model (GHM), Land Surface Model (LSM) and Dynamic Global Vegetation Model (DGVM).

Model	Model type	PET scheme	Runoff scheme	Capil. rise	Pref. flow	Soil layers	Total soil depth	Recharge definition	Calibration details
WaterGAP 2	GHM	Priestley-Taylor (Priestley and Taylor, 1972)	HBV (Bergström, 1976)	No	No	1	4 m	Fraction of runoff. In semi-arid, arid, or coarse soils, precipitation must exceed threshold for recharge to occur.	Calibrated against mean annual streamflow at 1319 GRDC stations
PCR-GLOBWB	GHM	Hamon (Hamon, 1963)	ARNO (Dumenil and Todini, 1992)	Yes	No	2	1.5 m	Net flux (percolation, capillary rise) from lowest soil layer to groundwater layer.	No calibration – adjustment of some parameters
MATSIRO	LSM	Monin-Obukhov (AET) (Kumiko et al., 2003)	TOPMODEL (Beven and Kirkby, 1979)	Yes	No	13	100 m	Net flux (gravitational drainage, capillary rise) from layer where water table sits.	No calibration – adjustment of parameters according to vegetation and soil properties
LPJmL	DGVM	Modified Priestley-Taylor (Gerten et al., 2004)	Bucket (Arnold et al., 1990)	No	No	6	13 m	Percolation from bottom soil later.	Calibrated against crop yields from FAO data
JULES-W1	LSM	Penman-Monteith (Best et al., 2011)	TOPMODEL (Beven and Kirkby, 1979)	No	No	4	3 m	Saturation excess water from the bottom soil layer.	Calibrated against biophysical processes
CWatM	GHM	Penman-Monteith (Allen et al., 1998)	ARNO (Dumenil and Todini, 1992)	Yes	Yes	3	2 m	Net flux (percolation, preferential flow, capillary rise) from lowest soil layer to groundwater store.	Calibrated against monthly and daily streamflow data in 12 catchments
H08	GHM	Bulk transfer coefficient (Hanasaki et al., 2008)	Leaky bucket (Manabe, 1969)	No	No	1	1 m	Same as WaterGAP 2.	No calibration
CLM 4.5	LSM	Monin-Obukhov (AET) (Oleson et al., 2013)	TOPMODEL (Beven and Kirkby, 1979)	Yes -via soil matrix potential	No	15	42.1 m	Net flux (percolation, capillary rise) to groundwater layer. Applies Darcy's law across in the groundwater layer.	No calibration – adjustment of parameters according to vegetation and soil properties

and because of its focus on Africa, we selected this database above other meta-datasets (Moeck et al., 2020; Mohan et al., 2018).

We further screened the dataset derived previously to only include georeferenced and time-stamped findings between 1979 and 2005, enabling a comparison to global models. This meant that we ultimately use 124 (out of originally 134) ground-based estimates of annual recharge across Africa. Our analysis of recharge ratio uses fewer data points, as only 106 of these estimates report corresponding mean annual precipitation values. Spatially, 28 of these estimates reflect recharge rates over spatial scales <100 km², a further 39, 29, and 28 are for spatial scales of 100–2500 km², 2500–62,500 km² and >62,500 km², respectively.

2.3. Random forests and predictor variables

We used random forests to predict the simulated long-term annual recharge and recharge ratio estimates using environmental attributes (Table A1 in the appendix) to assess environmental controls on modelled outputs. A random forest is a supervised machine learning algorithm that combines multiple trees to produce an ensemble of predictions (Breiman et al., 1984; Breiman, 2001), which link predictor variables (environmental attributes) to a response (global model outputs). As discussed by Addor et al. (2018), the advantages of random forests include no relationship assumptions, the allowance of non-linear relationships between multiple predictors, reduced risk of overfitting compared to individual regression trees, and computational efficiency. Each regression tree in the ensemble model is trained on observations (model grid cells) randomly selected with replacement from a sub-sample of 70 % of the total observations ('in-bag' observations). In each forest, we use 250 trees, each of which can make a maximum of 400 decision splits. Greater numbers of trees or splits did not improve the accuracy of the predictions.

To interpret the random forest models, we follow an approach taken by Addor et al. (2018) and group predictor variables according to climate, landcover, topography, soils and geology. Independent random forest models are then developed for each predictor group when estimating recharge outputs from the global models. Determining the R^2 values for the 'out of bag' (i.e., not training data) predictions from each ensemble model

then allows us to see how much of the variability in global model outputs can be explained by either climate, landcover, topography, soils, or geology; as well as jointly using all variables. Information about the predictor variables/attributes in each predictor group and the datasets used to characterise these variables can be found in Table A1 in the appendix.

3. Results and discussion

We organise the following results and discussion according to the three research questions presented at the end of the introduction. For each question, the results are presented and then immediately discussed.

3.1. Where do recharge estimates from global-scale models agree or disagree?

We assess the agreement of annual recharge and recharge ratio estimates from the eight global models by investigating the standard deviations (using absolute values) and coefficients of variation (also called relative standard deviation using normalized values) across the model outputs (Fig. 1). In absolute values, we find that annual recharge estimates from the eight models disagree considerably in wetter regions of Africa (Fig. 1. a). Disagreements in annual recharge estimates are greatest in Central Africa, the Ethiopian highlands, Madagascar, and along the west African coastline, with 24 % of pixels having standard deviations above 100 mm/year. In contrast, disagreements in annual recharge estimates are <10 mm/year for 38 % of pixels, predominantly extending across dry regions of the continent such as the Sahara, Southern Africa, and the Horn of Africa. Though, even in some very dry regions, we find places where annual recharge estimates disagree, predominantly along the river Nile and irrigated agriculture regions of the Sahara. In these locations, CLM 4.5, CWatM, H08, LPJmL, MATSIRO all estimate recharge ratios greater than one (Figs. S1 & S2), likely reflecting modelled transmission losses or irrigation return flows.

Recharge ratio estimates disagree across both wet and dry regions (Fig. 1.b). They disagree across much of the Sahara, with standard deviations >0.1. However, most of this divergence derives from only one model, PCR-GLOBWB, which estimates recharge ratios >0.2 for nearly all
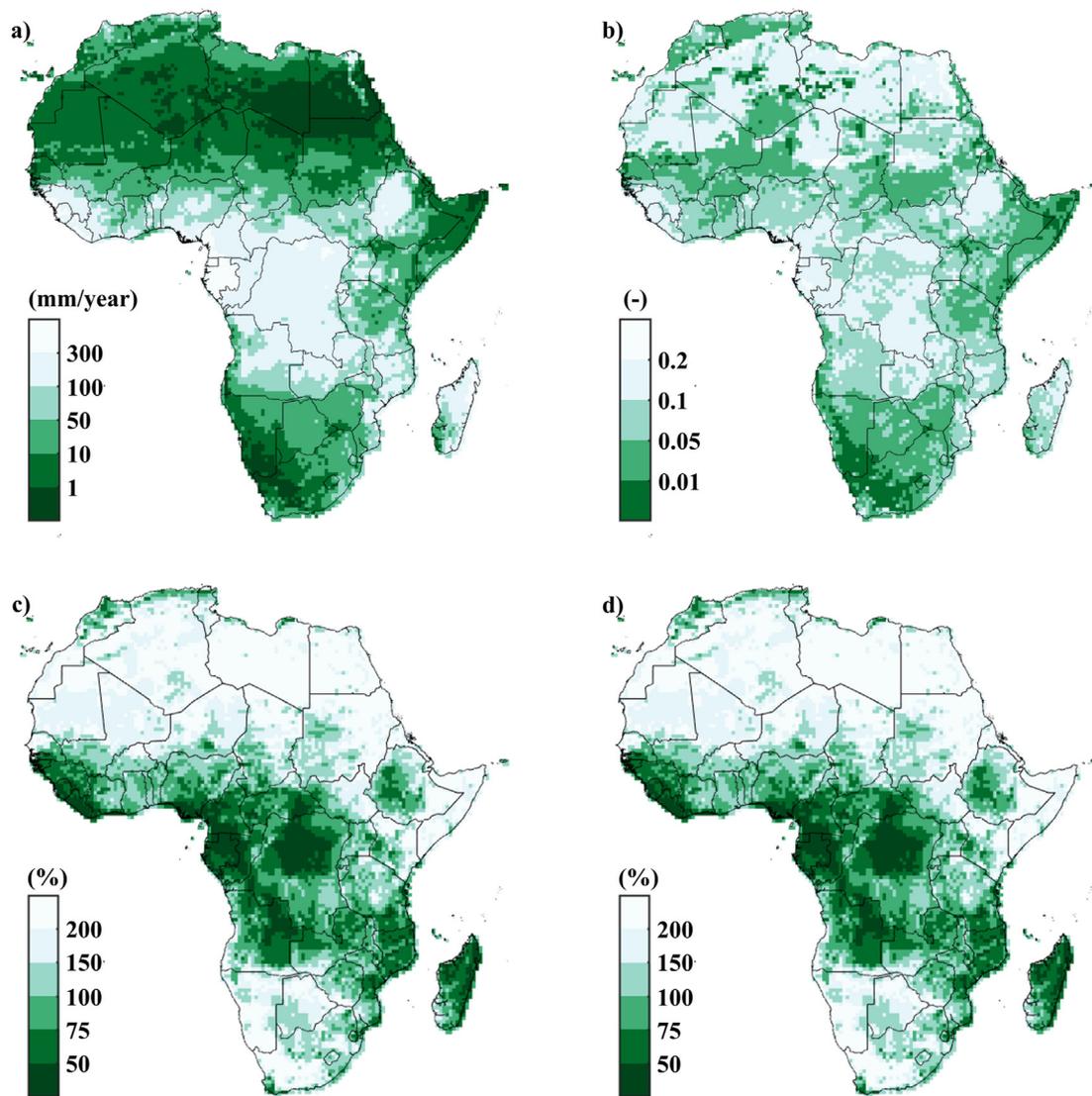
Fig. 1. Maps show joint summary statistics of the variability in recharge estimates across Africa from the 8 global models listed in Table 1 and discussed in detail in Reinecke et al. (2021). (a) the standard deviation of long-term average annual recharge rates (mm/year); (b) the standard deviation of long-term average recharge ratios (−). (c) the coefficient of variation of long-term average annual recharge rates (%); (d) the coefficient of variation of long-term average recharge ratios (%). The colouring scheme we have selected is intended to highlight where our understanding of recharge varies or converges. Lighter colours highlight where model estimates diverge and hence where our understanding of recharge processes is incoherent (Wagener et al., 2021). Whereas darker colours show where models agree, though this does not mean they are correct.

Northern Africa (Fig. S2). The other seven models estimate recharge ratios below 0.05 for most of the Sahara. Modelled recharge ratio estimates also disagree throughout Central Africa and the Ethiopian Highlands, such that 64 % of pixels have a standard deviation >0.1 for recharge ratio. For 32 % (5 %) of the pixels, standard deviations in recharge ratio estimates from the eight models are below 0.05 (0.01), primarily distributed across the Sahel, Southern Africa, and the Horn of Africa.

The picture changes if we compare estimates of mean annual recharge (Fig. 1.c) and recharge ratio (Fig. 1.d) relative to their ensemble mean. The relative disagreement between models is high where estimated recharge rates are low, and vice versa. Therefore, model differences tend to be either high in magnitude (large standard deviation, small coefficient of variation) or high relative to estimated recharge (small standard deviation, large coefficient of variation). Identifying regions where model disagreement is high relative to recharge flux is important, especially if a later model use lies in the estimation of climate change impacts (e.g., Hartmann et al., 2017). Here, the coefficient of variation (i.e., relative standard deviation) between modelled estimates is >100 %

(200 %) for 66 % (29 %) of pixels in Africa, mostly occurring in Northern and Southern Africa as well as the Horn of Africa. Whilst only 6 % of pixels have a coefficient of variation <50 %. Furthermore, as relative differences in modelled annual recharge estimates are equal to relative differences in modelled recharge ratio estimate, they show the same spatial patterns (Fig. 1.c & d). Hence highlighting how differences in annual recharge estimates are caused by differences in precipitation-recharge conversion rates, which could be attributed to varying model structures, parameterisation schemes or datasets used to characterise the land surface. Further analysis investigating model differences at the global scale are included in the supplemental information (Figs. S3 & S4), along with a discussion about previous model comparisons for groundwater recharge in other parts of the world.

Fig. 2 shows that both absolute and relative differences in groundwater recharge estimates vary strongly with long-term mean annual P-PET. In disagreement with previous empirical studies (MacDonald et al., 2021; West et al., 2022), recharge ratio estimates from PCR-GLOBWB and CWatM are surprisingly high (>0.2) across desert regions where P-PET is low, leading
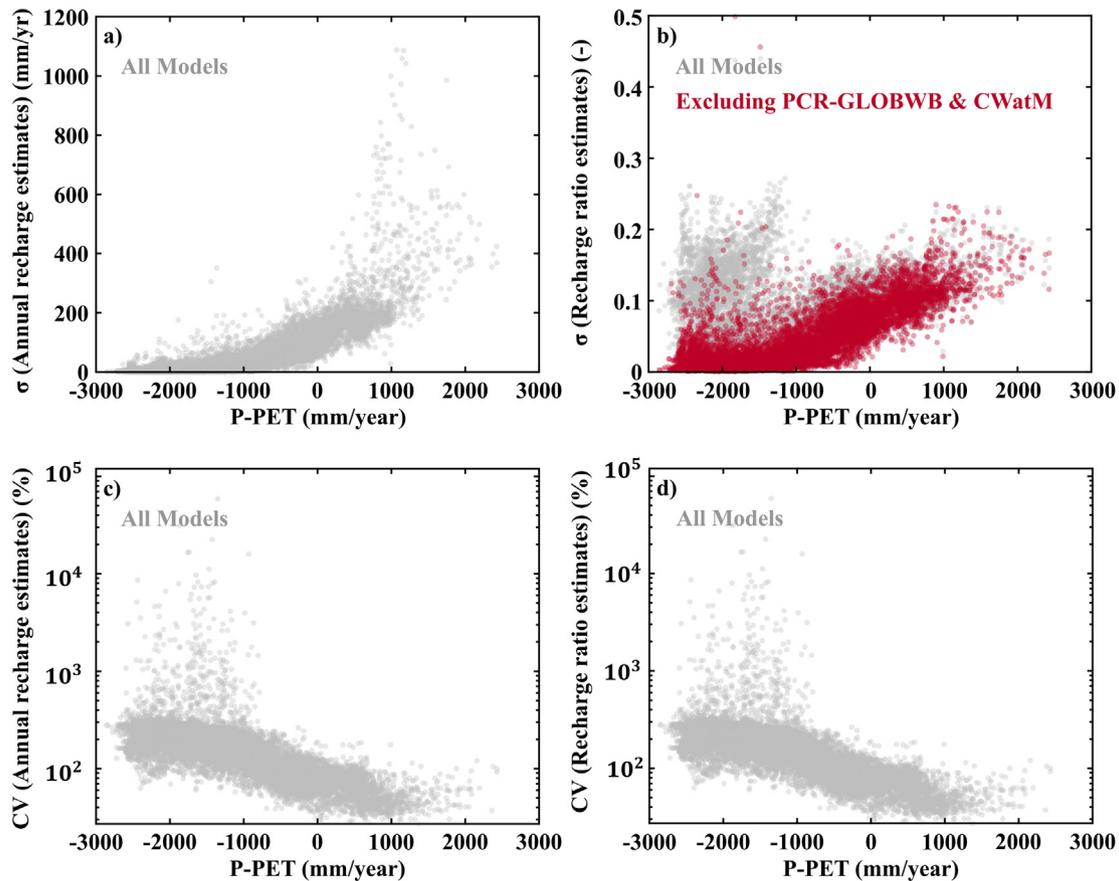
Fig. 2. Further analysis of the recharge estimates for Africa from the global models listed in Table 1 and shown in Fig. 1. (a) long-term average P-PET (mm/year) against the standard deviation of long-term average annual recharge rates (mm/year); (b) long-term average P-PET (mm/year) against the standard deviation of long-term average recharge ratios; (c) long-term average P-PET (mm/year) against the coefficient of variation of long-term average annual recharge rate; (d) long-term average P-PET (mm/year) against the coefficient of variation of long-term average recharge ratios.

to large absolute model disagreements for recharge ratio throughout dry and wet regions (Fig. 1.b & Fig. S2). The reasons for this deviation are not obvious.

3.2. Which environmental controls describe the recharge patterns produced by the different models?

We use random forest models to investigate which environmental controls, as identified from datasets, with global coverage, describe the spatial patterns evident in the recharge outputs from global-scale models. The logic is that recharge outputs from global-scale models should reflect similar process controls to what is being identified in continental to global-scale empirical studies found in the literature. If global-scale models show very different controls to what is being found through empirical studies, then this may indicate that the model structures of global-scale models do not adequately represent dominant recharge controls.

Climate variables explain most of the spatial variability in annual recharge and recharge ratio estimates from all eight models, with landcover and soil properties also showing some explanatory power (Fig. 3). The static descriptors of climate, landcover, and soils used in the random forest models explain at least 80 % of the modelled spatial variability of annual recharge and at least 70 % of the spatial variability of recharge ratio. Climate alone explains between 67 % and 89 % of the spatial variability in long-term annual recharge estimates from the eight global-scale models. For five of the models, climate also explains >70 % of the spatial variability in recharge ratio estimates. For the remaining three models (PCR-GLOBWB, LPJmL, CWaTM), just over 50 % of the spatial variability of recharge ratio can be explained by climate alone (53 %, 65 %, 51 %). Landcover attributes are second in explaining the spatial variability in recharge outputs from all

eight global models. These attributes explain on average 61 % (56 %) of annual recharge (recharge ratio) estimates, whilst soil properties explain on average 33 % (34 %) of their spatial variability.

Considering the interactions between climate, landcover, topography, soils and geology allows us to explain only small additional proportions of the spatial variability in both annual recharge and recharge ratio estimates from the global models. For annual recharge (recharge ratio), the additional explanatory power when considering interactions between all predictor variables in contrast to climate controls alone, is on average 10 % (13 %) across each of the global-scale models. This might be expected as the co-evolution of climate, landcover and soils causes these properties to co-vary in space and form large-scale landscape patterns (Pelletier et al., 2013; Troch et al., 2013). Therefore, using climate variables alone to predict modelled recharge outputs implicitly considers some information about landcover and soils.

The importance of climate controls on global model outputs is consistent with the findings of MacDonald et al. (2021) for ground-based recharge estimates. Their regression showed that ground-based annual recharge estimates vary throughout Africa according to mean annual precipitation, which explains 82 % of this variability. Yet in contrast to our global model findings, they do not find that including additional variables beyond climate improve their prediction of ground-based recharge estimates across Africa, though they suggested that it could be important locally. The low level of predictability of climate controls on recharge estimates from PCR-GLOBWB, LPJmL and CWaTM, shows that these models are inconsistent with the controls found for ground-based estimates in Africa (MacDonald et al., 2021). Other empirical studies, analysing annual recharge rates globally, do however highlight the importance of vegetation and soils for partitioning precipitation to recharge (Kim and Jackson,
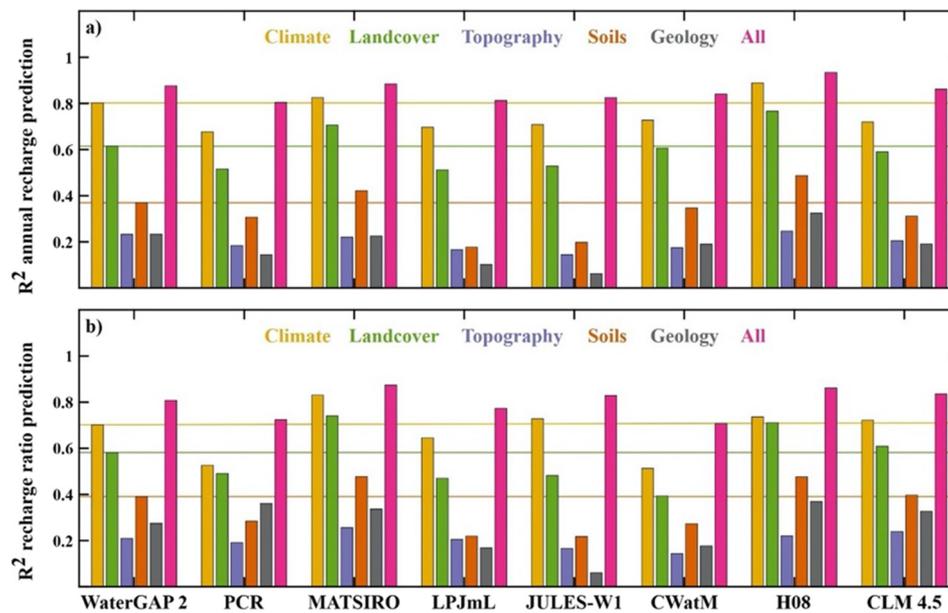
Fig. 3. Coefficient of determination when predicting global model estimates for Africa of (a) long-term average annual recharge and (b) recharge ratio. Out of bag predictions using random forest models were performed for individual predictor groups (climate, landcover, topography, soils, and geology) as well as by considering the interaction between all predictor attributes (all). Random forest models are an ensemble of 250 regression trees each with a maximum of 400 decision splits. When predicting recharge ratios, we excluded pixels with estimates greater than one as their inclusion led to very low coefficient of determination scores when using all predictor attributes. Percentage of pixels excluded when predicting recharge ratio estimates were <0.5 % for all models. Horizontal lines highlight the three models which consistently reflect the strongest levels of climate, vegetation and soil controls on their annual recharge and recharge ratio estimates.

2012; Mohan et al., 2018). The differences between these global analyses and MacDonald et al. (2021) could highlight particularly strong climate controls on the spatial variability of recharge in Africa. Whereas precipitation gradients in other regions may not be as great as those in Africa, which may explain why landcover and soil properties have helped to explain the variability of recharge globally. It may also be attributed to spatial patterns becoming more emergent in larger (>600 datapoints) datasets covering more environments or more rigorous data quality assurance by MacDonald et al. (2021). Quality assurance procedures by MacDonald et al. (2021) led to the removal of 182 (from an initial 316) data points, which may have reduced the variability in recharge estimates and therefore produced a clearer climatic control.

3.3. How do model estimates compare to ground-based estimates compiled from the literature?

Ground-based recharge estimates are distributed unevenly across the African continent and are predominantly located in dryland landscapes whilst wet tropical landscapes are underrepresented (Fig. 4). For desert, dryland, wet tropical, and wet tropical forest landscapes we have 21 (16), 71 (61), 29 (26), and 3 (3) ground-based estimates of annual recharge (recharge ratio), respectively. We compare these estimates to those from the global-scale models for the relevant period. A description of how we delineated the four landscapes is provided by West et al. (2022).

Discrepancies between global-scale model and ground-based estimates of annual recharge and recharge ratio are larger in wetter than in drier recharge landscapes, though relative discrepancies are similar across the different environmental settings (Fig. 4). The median magnitude of discrepancies in annual recharge (recharge ratio) in desert, dryland, wet tropical and wet tropical forest landscapes are 0.4 mm/year (0.007), 4.26 mm/year (0.01), 31.7 mm/year (0.02) and 105 mm/year (0.06), respectively when looking across all eight global models. In contrast, the median magnitude of relative discrepancies in annual recharge (recharge ratio) in desert, dryland, wet tropical and wet tropical forest landscapes are 23 % (30 %), 29 % (30 %), 29 % (22 %) and 12 % (30 %), respectively. Furthermore, relative discrepancies in annual recharge estimates are often similar if not identical to relative discrepancies in recharge ratio. For each

model, linear correlations between the relative discrepancies in annual recharge and recharge ratio vary between 0.96 and 0.99 (Fig. S5).

Landscape specific discrepancies between global models and ground-based recharge estimates are more noticeable for several global models. CLM 4.5 and CWatM show the greatest overpredictions in wet tropical forest landscapes, while H08, JULES, and WaterGAP 2 significantly underpredict in this domain. PCR-GLOBWB and CWatM show the largest overpredictions in terms of recharge ratios in desert landscapes. Climate controls on recharge ratio estimates from these two models were less dominant than for the other models (Fig. 4), which together with generally large overestimations in recharge ratio suggests that they do not represent recharge controls in desert regions adequately. CLM 4.5 also displays larger overpredictions in recharge ratio in wet tropical forest landscapes. It is also interesting to note that in some model and landscape combinations, discrepancies to ground-based estimates show very little bias but a lot of variability, whilst in others the bias can be high but with much less variability. This is likely due to the modelling decisions made by each of the model development groups, such as model structure, calibration procedures and dataset selection (Telteu et al., 2021). However diagnosing which of these decisions are responsible for our findings is non-trivial and will likely require further discussion and assessment between and by model developers (Gudmundsson et al., 2012). Additionally, comparing global model outputs to ground-based information also raises questions about how comparable the model spatial resolution is to the representative area of the ground-based data. However, we did not find a clear relationship between the magnitude of global model discrepancies and the varying spatial scales of ground-based recharge estimates (Fig. S6).

We find that model similarity to ground-based estimates varies considerably and inconsistently throughout the different Recharge Landscapes of Africa (Fig. 4). Even though six models predominantly underestimate mean annual groundwater recharge rates, this is not a consistent pattern, with over-estimates present for most models and in each landscape (Fig. 5). JULES-W1 appears to be the exception to this, as 87 % (92 %) of its annual recharge (recharge ratio) estimates are below ground-based findings. Across all eight models, we found that 56 % (55 %) of annual recharge (recharge ratio) discrepancies are underestimates, 18 % (18 %) are overestimates, and 26 % (27 %) fell within the uncertainty range of the ground-
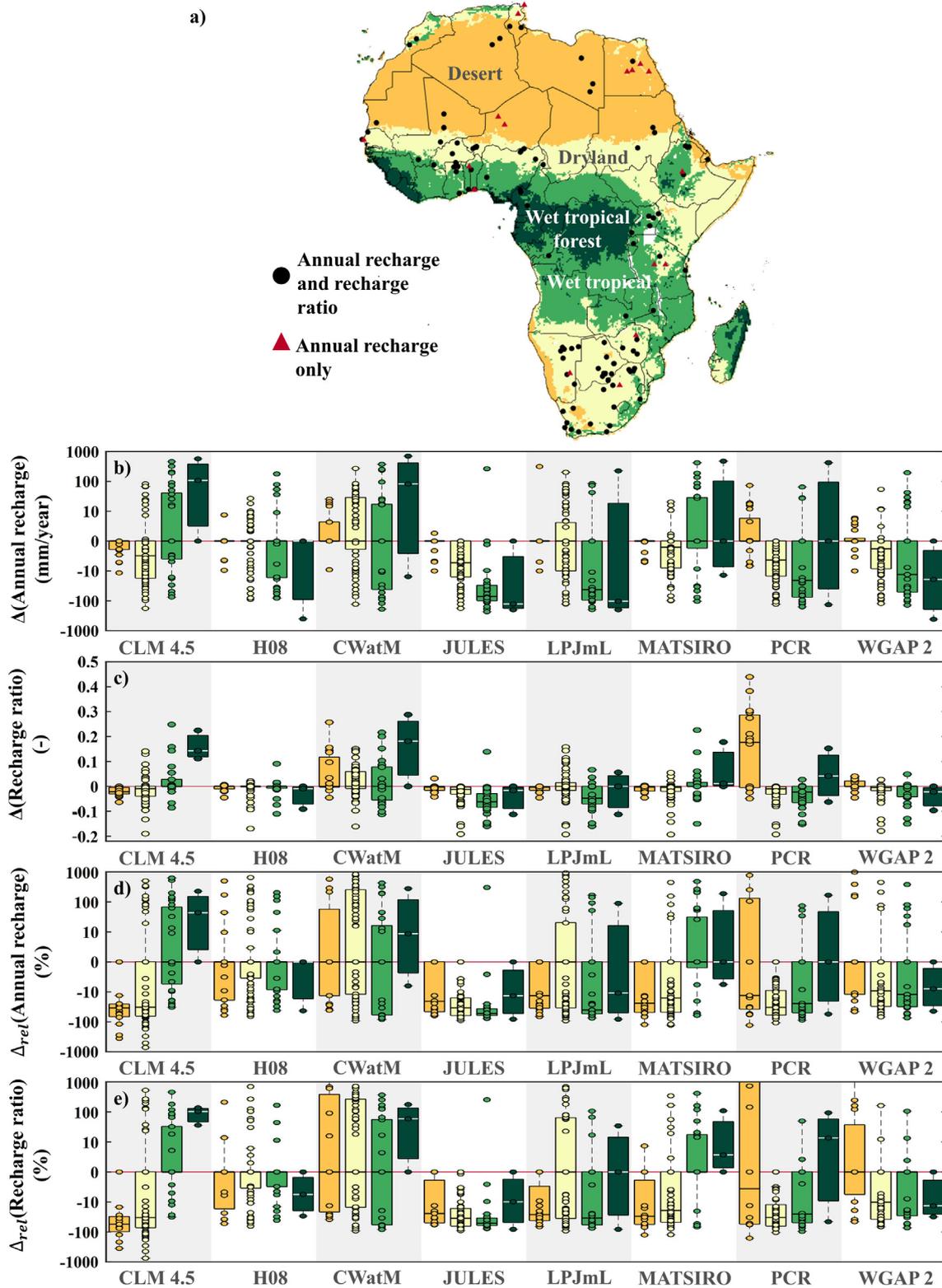
Fig. 4. Map of Africa organised into four recharge landscapes (West et al., 2022) with the distribution of ground-based annual recharge and recharge ratio (annual recharge/ annual precipitation) estimates superimposed. Discrepancy between global model and ground-based estimates of (b) annual recharge (mm/year) and (c) recharge ratio (−) organised according to these recharge landscapes. Discrepancy is defined as global model estimate - ground-based estimate. Relative discrepancy between global model and ground-based estimates of (d) annual recharge (%) and (e) recharge ratio (%) organised according to these recharge landscapes. Relative discrepancy is defined as 100 X (global model estimate - ground-based estimate)/ground-based estimate. Scatter points show the individual relative discrepancies and boxplots show the inter-quartile range of relative discrepancies in each landscape. Scatter points show the individual discrepancies or relative discrepancies, and boxplots show their inter-quartile range in each landscape. Some ground-based estimates have uncertainty ranges. If global model estimates fall within this range the discrepancy and relative discrepancy is zero.
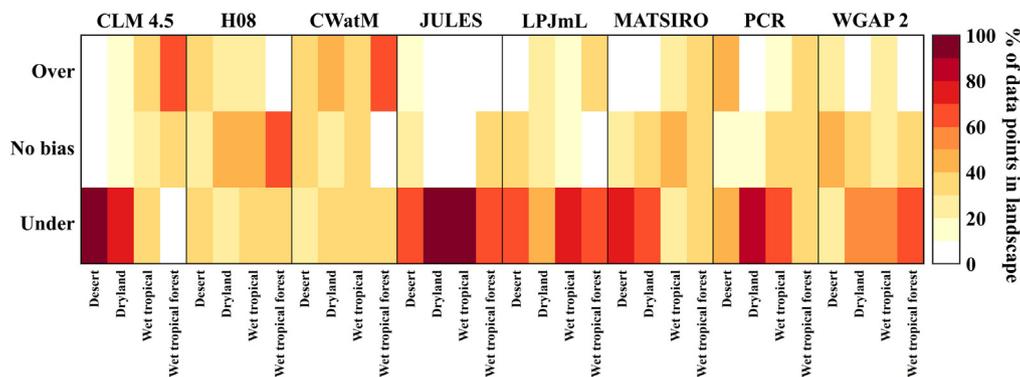
Fig. 5. The percentage of data points in a given recharge landscape where global-scale models either over or underestimate ground-based annual recharge estimates or where there is no bias (i.e., the global-scale model estimate is within the uncertainty range of the ground-based estimate).

based estimate. Similar over and underestimation statistics for recharge rates and recharge ratios, reflect that overestimation (underestimation) in annual recharge mostly corresponds to overestimation (underestimation) in recharge ratio. The tendency to underestimate recharge rates is particularly noticeable in dryland recharge landscapes, where 60 % of the global model estimates are below ground-based estimates. Though this varies from 28 % to 92 % across each of the models.

Global models which reflect the strongest levels of climate controls on their recharge outputs agree with ground-based estimates more than the other models. Recharge outputs from H08, MATSIRO, and WaterGAP 2, consistently reflect more climate controls than the other models and show lower average discrepancies. The median absolute discrepancies in annual recharge (recharge ratio) for these three models are 0.36 mm/year (0), 1.56 mm/year (0.006) and 2.9 mm/year (0.006), respectively, in contrast to between 5 mm/year and 10 mm/year (0.016 and 0.03). This agrees with previous empirical analyses of ground-based recharge estimates by MacDonald et al. (2021) and West et al. (2022) who both found climate to be the dominant controls on spatial variability in Africa. Moreover, Zaherpour et al. (2018) found these three models performed better than others when estimating mean annual runoff at 40 catchments distributed throughout the world. Hence potentially suggesting the greater plausibility of these model structures for long-term hydrological partitioning (Troch et al., 2013). Median absolute discrepancies for each model are generally low as >70 % of datapoints are in desert or dryland landscapes. Interestingly, the median absolute discrepancy in annual recharge (recharge ratio) estimates for the ensemble mean of the models is 1.11 mm/year (0.005), which is similar to H08, MATSIRO and WaterGAP 2.

Based on these findings, we suggest predicting groundwater recharge across continental scales for research or operational management should not rely on one specific model (Fletcher et al., 2019). No model consistently compares well with ground-based estimates in a given landscape and therefore we cannot rely upon individual global models to estimate recharge. If groundwater governance in wet landscapes depends on recharge estimates from one global-scale model, this could potentially lead to large over or under-utilization of groundwater resources, as absolute discrepancies between global models and ground-based estimates are often large in these settings. Although absolute discrepancies between global model and ground-based estimates are often small in dry and desert landscapes, relative to the ground-based estimate and to the total available water resources of the region, these discrepancies are often large. It is perhaps not surprising that global models are unreliable in wet tropical regions as so few studies have investigated recharge processes in these settings (Mohan et al., 2018; Moeck et al., 2020; MacDonald et al., 2021). Nonetheless, recharge studies are much more common in dry settings, and this is still not leading towards improved recharge estimation in relative terms, by global-scale models. Hence highlighting the need to advance recharge estimation by global-scale models across a wide range of environmental settings. Interestingly, the performance in absolute terms of the ensemble mean of the global-scale models is like that of the individually better performing models

(H08, MATSIRO and WaterGAP 2), for both annual recharge and recharge ratio. This is perhaps not surprising as these models reflect the greatest level of climate controls, and by taking the ensemble mean across all the global-scale models, we essentially de-emphasize how individual models partition water at the land surface and instead emphasize the climate forcing.

However, it is still unclear how we can improve recharge estimation by global-scale models as their discrepancies to ground-based estimates are inconsistent, with both over and under-estimates by models in a given landscape. If biases were systematically either positive or negative this could help guide further diagnostic model evaluation considering other hydrologic fluxes (Niraula et al., 2017), but this is not the case. Previous comparison studies have shown that the development of consistent modelling protocols that harmonise model inputs, parameter estimation, initial conditions etc. is very difficult even for simple models (Ceola et al., 2015). In the long-run, modular modelling approaches will enable a better implementation of different model structures for global hydrologic models. Such strategies have been widely used for lumped hydrologic models at the catchment scale (e.g. Clark et al., 2008; Knoben et al., 2019; Leavesley et al., 2002; Wagener et al., 2001). However, developing such frameworks for distributed global-scale models will not be straightforward and would best be done as a large-scale collaboration between different modelling groups. Successful implementation would exceed the levels of diagnostic analysis currently possible in existing model intercomparison projects, such as ISIMIP. Further still, if this modular framework facilitated spatially variable model structures, this would allow the simultaneous implementation of landscape-specific models (Hartmann et al., 2017; Quichimbo et al., 2021), which may be more plausible than models which uniformly apply the same model structure over entire continents or globally. Developing a sensible separation of the landscape for specific models and capturing the expert understanding of how these different systems function is an interesting current challenge (Gleeson et al., 2021). Using expert knowledge to develop more plausible models is particularly important, as even a direct calibration of models to available data may not improve recharge estimation due to sparsity of recharge data (see Section 2 of the supplemental information for further discussion).

4. Conclusions

We set out to examine where and how global-scale model estimates of groundwater recharge agree or disagree with one another and with ground-based estimates across the African continent. We did so using the outputs of eight global models that were previously run within the ISIMIP framework (Frieler et al., 2017; Reinecke et al., 2021) and over 100 ground-based estimates of recharge compiled from the literature (MacDonald et al., 2021).

We found that global-scale model estimates of long-term mean annual recharge rates and recharge ratio disagree significantly throughout much of Africa. In absolute terms (using standard deviation), models disagree

more in wet tropical regions and agree more in dry regions, whilst in relative terms (using coefficient of variation) the opposite is true. However, absolute model disagreement for recharge ratios is also high throughout the Sahara Desert, though this is mostly attributed to surprisingly high estimates by PCR-GLOBWB and CWatM in this region. When investigating controls on global-scale model recharge outputs for Africa using Random Forests, we found that climate controls on average explain 75 % and 68 % of the spatial variability of annual recharge and recharge ratio estimates, respectively. However, climate controls only explained approximately 60 % of the spatial variability in recharge estimates from PCR-GLOBWB and CWatM, which is significantly less than the 82 % reported by MacDonald et al. (2021) in their empirical analysis. H08, MATSIRO and WaterGAP 2, reflected the strongest level of climate control on their recharge outputs and show the greatest similarity to ground-based estimates. Using the ensemble mean of all eight global-scale models performed similarly to these three best performing models.

Our work adds further evidence to previous studies which showed that the robustness of global hydrologic model simulations can vary considerably, suggesting that we should be aware of these problems when utilizing these model outputs for subsequent studies. Studies regularly use the output of such models (and often just one of them) as input in their follow-up analysis, only with the occasional caveat in the discussion section about potential limitations and robustness issues in the model predictions. Rather, highlighting where model robustness is low and where models deviate from our current perception of hydrologic systems are important strategies to guide future research towards areas of greatest knowledge gaps (Wagener et al., 2021).

CRediT authorship contribution statement

Charles West: Conceptualization, Methodology, Investigation, Formal Analysis, Writing Original Draft, Writing Review and Editing, Visualization.

Robert Reinecke: Resources, Writing Review and Editing. **Rafael Rosolem:** Conceptualization, Methodology, Writing Review and Editing, supervision. **Alan M MacDonald:** Conceptualization, Resources, Writing Review and Editing. **Mark O Cuthbert:** Conceptualization, Writing Review and Editing. **Thorsten Wagener:** Conceptualization, Methodology, Writing Review and Editing, supervision.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A

Table A1
List of environmental attributes used in analysis organised by predictor groups (i.e., climate, landcover, topography, soils, geology), along with information about the global datasets used to characterise attributes.

Attribute	Description	Units	Period	Data source	Reference
Climate attributes					
P-PET	Mean annual precipitation minus mean annual PET.	mm/year	1979–2005	1. MSWEP v1.2 (Precipitation)	
P-PET in season	Mean annual volume of precipitation in excess to PET in months considered in-season. A month is considered in-season when P exceeds PET.	mm/year	1979–2005	Spatial res.: 0.25° Temporal res.: Daily	1. (Beck et al., 2017)
σ(P-PET)	The standard deviation of monthly P-PET	mm/month	1979–2005		
P10	The average volume of rainfall per year on days with at least 10 mm of rain.	mm/year	1979–2005	2. CRU v4 (PET) Spatial res.: 0.5° Temporal res.: Monthly	2. (Harris et al., 2020)
Topography attributes					
Elevation	Elevation referenced to WGS84 EGM96 geoid	m	–		
Slope	Geodesic slope of the DEM using a 3 by 3 moving window.	degrees	–	HydroSHEDS Spatial res.: 15″	(Lehner et al., 2013)
HAND	Height Above Nearest Drainage. Height above nearest river pixel when following D8 flow directions.	m			
Landcover/use					
Kveg	Vegetation coefficient related to transpiration. Vegetation-specific annual values (Gordon et al., 2005) applied to a landcover classification. Mean value from 1992 to 2005.	–	1992–2005	ESA-CCI v2.0.7 Spatial res.: 300 m	(Defourny et al., 2017)
Wet	Extent of areas with LCCS class of 20, 160, 170 or 180	–	1992–2005	Temporal res.: Yearly	
Waterbodies	Extent of areas with LCCS class 210	–	1992–2005		
LAI	Mean leaf area index (based on 12 monthly means from 1981 to 2015)	–	1981–2015	GIMMS-LAI3g v2 Spatial res.: 0.25° Temporal res.: Monthly	(Mao and Yan, 2019)
Irrigation	Area equipped for irrigation multiplied by the fractional area actually irrigated.	km²	2005	Global Map of Irrigation Areas Spatial res.: 5′	(Siebert et al., 2013)

Table A1 (*continued*)

Attribute	Description	Units	Period	Data source	Reference
Soil attributes					
Sand	Proportion (by weight) of sand particles (>0.05 mm) in the fine earth fraction of the upper 2 m of the subsurface. Take the depth weighted harmonic mean across intervals of 0-5 cm, 5-15 cm, 15-30 cm, 30 cm–60 cm, 60-100 cm, 100-200 cm.	%	–		
Clay	Proportion (by weight) of clay particles (<0.002 mm) in fine earth fraction of the upper 2 m of the subsurface. Take the depth weighted harmonic mean across intervals of 0-5 cm, 5-15 cm, 15-30 cm, 30 cm–60 cm, 60-100 cm, 100-200 cm.	%	–	SoilGrids250m Spatial res.: 250 m	(Hengl et al., 2017)
Sand/(Clay + Silt)	The ratio of sand to clay and silt in the top 2 m of the soil profile.	–	–		
Geology attributes					
Depth to Bedrock	Depth to bedrock	m	–	Soil grids 1 km Spatial res.: 1 km	(Hengl et al., 2014)
Depth to bedrock 2	Average soil and sedimentary deposit thickness. Maximum of 50 m.	m	–	Gridded thickness of soil, regolith and sedimentary deposit layers Spatial res.: 30″	(Pelletier et al., 2016)
Karst	Extent of carbonate rock outcrop areas.	–	–	World Map of Carbonate Rock Outcrops V3.0	(Williams and Ford, 2006)

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.159765.

References

Addor, N., Nearing, G., Prieto, C., Newman, A.J., Le Vine, N., Clark, M.P., 2018. A ranking of hydrological signatures based on their predictability in space. Water Resour. Res. 8792–8812.

Aeschbach-Hertig, W., Gleeson, T., 2012. Regional strategies for the accelerating global problem of groundwater depletion. Nat. Geosci. 5 (12), 853–861.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - guidelines for computing crop water requirements - FAO irrigation and drainage paper 56 Rome.

Arnold, J.G., Wlliams, J.R., Nicks, A.D., Simmons, N.B., 1990. SWRRB - A Basin Scale Simulation Model for Soil and Water Resources Management.

Beck, H.E., Van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., De Roo, A., 2017. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615.

Bergström, S., 1976. Development and application of a conceptual runoff model for scandinavian catchments. Smhi RHO 7 (November), 134.

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., Harding.R.J., 2011. The joint UK land environment simulator (JULES), model description – part 1: energy and water fluxes. Geosci. Model Dev. 4 (3), 677–699.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24 (1), 43–69.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. 1st ed. Chapman and Hall/CRC.

Calow, R.C., MacDonald, A.M., Nicol, A.L., Robins, N.S., 2010. Ground water security and drought in Africa: linking availability, access, and demand. Ground Water 48, 246–256.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A., Wagener, T., 2015. Virtual laboratories: new opportunities for collaborative water science. Hydrol. Earth Syst. Sci. 19 (4), 2101–2117.

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T., Hay, L.E., 2008. Framework for understanding structural errors (FUSE): a modular framework to diagnose differences between hydrological models. Water Resour. Res. 44 (12).

Collins, W.J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Hinton, T., Jones, C.D., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Totterdell, I., Woodward, S., Reichler, T., Kim, J., 2008. Evaluation of HadGEM2 model. Meteorological Office Hadley Centre, Technical Note 74, 47 pp.

Cuthbert, M.O., Gleeson, T., Moosdorf, N., Befus, K.M., Schneider, A., Hartmann, J., Lehner, B., 2019. Global patterns and dynamics of Climate–groundwater interactions. Nat. Clim. Chang. 9 (2), 137–141.

Dalin, C., Wada, Y., Kastner, T., Puma, M.J., 2017. Groundwater depletion embedded in international food trade. Nature 543 (7647), 700–704.

de Graaf, I.E.M., Sutanudjaja, E.H., Van Beek, L.P.H., Bierkens, M.F.P., 2015. A high-resolution global-scale groundwater model. Hydrol. Earth Syst. Sci. 19 (2), 823–837.

de Graaf, I.E.M., van Beek, R.L.P.H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E.H., Bierkens, M.F.P., 2017. A global-scale two-layer transient groundwater model: development and application to groundwater depletion. Adv. Water Resour. 102, 53–67.

de Graaf, I.E.M., Gleeson, T., van Beek, R.L.P.H., Sutanudjaja, E.H., Bierkens, M.F.P., 2019. Environmental flow limits to global groundwater pumping. Nature 574 (7776), 90–94.

Defourny, P., Bontemps, S., Lamarche, C., Brockmann, C., Boettcher, M., Wevers, J., Kirches, G., Santoro, M., 2017. Land Cover CCI Product User Guide - Version 2.0. ESA 1–105.

Döll, P., 2009. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. Environ. Res. Lett. 4 (3), 35006.

Döll, P., Fiedler, K., 2008. Global-scale modeling of groundwater recharge. Hydrol. Earth Syst. Sci. 12 (3), 863–885.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F.T., Eicker, A., 2014. Global-scale assessment of groundwater depletion and related groundwater abstractions: combining hydrological modeling with information from well observations and GRACE satellites. Water Resour. Res. 50 (7), 5698–5720.

Dumenil, L., Todini, E., 1992. A rainfall-runoff scheme for use in the Hamburg climate model. Adv. Theor. Hydrol. 129–157.

Fletcher, S., Lickley, M., Strzepek, K., 2019. Learning about climate change uncertainty enables flexible water infrastructure planning. Nat. Commun. 10 (1).

Frieler, K., Betts, R., Burke, E., Ciais, P., Denvil, S., Deryng, D., Ebi, T., Emanuel, K., Elliott, J., Galbraith, E., Gosling, S., Halladay, K., Hattermann, F., Hickler, T., Hinkel, J., Huber, V., Jones, C., Krysanova, V., Lange, S., Lotze, H., Lotze-Campen, H., Mengel, M., Mouratiadou, I., Müller Schmied, H., Ostberg, S., Piontek, F., Popp, A., Reyer, C., Schewe, J., Stevanovic, M., Suzuki, T., Thonicke, K., Tian, H., Tittensor, D., Vautard, R., van Vliet, M., Warszawski, L., Zhao, Fang, 2017. Assessing the impacts of 1.5 °C global warming – simulation protocol of the inter-sectoral impact model intercomparison project (ISIMIP2b). Geosci. Model Dev. 10, 4321–4345.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., Sitch, S., 2004. Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model. J. Hydrol. 286 (1–4), 249–270.

Gleeson, T., Wada, Y., Bierkens, M.F.P., Van Beek, L.P.H., 2012. Water balance of global aquifers revealed by groundwater footprint. Nature 488 (7410), 197–200.

Gleeson, T., Befus, K.M., Jasechko, S., Luijendijk, E., Cardenas, M.B., 2016. The global volume and distribution of modern groundwater. Nat. Geosci. 9, 161–167.

Gleeson, T., Cuthbert, M., Ferguson, G., Perrone, D., 2020. Global groundwater sustainability, resources, and Systems in the Anthropocene. Annu. Rev. Earth Planet. Sci. 48, 431–463.

Gleeson, T., Wagener, T., Doll, P., Zipper, S.C., West, C., Wada, Y., Taylor, R.G., Scanlon, B., Rosolem, R., Rahman, M., Oshinlaja, N., Maxwell, R., Lo, M.-H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J.S., Ducharne, A., de Graaf, I.E.M., Cuthbert, M.O., Condon, L.E., Bresciani, E., Bierkens, M.F.P., 2021. GMD perspective: the quest to improve the evaluation of groundwater representation in continental-to global-scale models. Geosci. Model Dev. 14 (12), 7545–7571.

Gordon, L.J., Steffen, W., Jonsson, B.F., Folke, C., Falkenmark, M., Johannessen, A., 2005. Human modification of global water vapor flows from the land surface. Proc. Natl. Acad. Sci. 102 (21), 7612–7617.

Gudmundsson, L., Wagener, T., Tallaksen, L.M., Engeland, K., 2012. Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. Water Resour. Res. 48 (11), W11504.

Hamon, W.R., 1963. Computation of direct runoff amounts from storm rainfall. Int. Assoc. Sci. Hydrol. 63, 52–62.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., 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 (4), 1007–1025.

Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientific Data 7 (1), 1–18.

Hartmann, A., Gleeson, T., Wada, Y., Wagener, T., 2017. Enhanced Groundwater Recharge Rates and Altered Recharge Sensitivity to Climate Variability through Subsurface Heterogeneity. Proc. Natl. Acad. Sci. 114 (11), 2842–2847.

Hengl, T., Mendes De Jesus, J., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Ruiperez Gonzalez, M.,

2014. SoilGrids1km - global soil information based on automated mapping. PLoS ONE 9 (8), e105992.

Hengl, T., Mendes De Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12 (2), e0169748.

Jasechko, S., Birks, S.J., Gleeson, T., Wada, Y., Fawcett, P.J., Sharp, Z.D., McDonnell, J.J., Welker, J.M., 2014. The pronounced seasonality of global groundwater recharge. Water Resour. Res. 50 (11), 8845–8867.

Kim, J.H., Jackson, R.B., 2012. A global analysis of groundwater recharge for vegetation, climate, and soils. Vadose Zone J. 11 (1).

Knoben, W.J.M., Freer, J.E., Fowler, K.J.A., Peel, M.C., Woods, R.A., 2019. Modular assessment of rainfall-runoff models toolbox (MARRMoT) v1.2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. Geosci. Model Dev. 12 (6), 2463–2480.

Kumiko, T., Emori, S., Watanabe, T., 2003. Development of the minimal advanced treatments of surface interaction and runoff. Glob. Planet. Chang. 38 (1–2), 209–222.

Lange, S., 2018. Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. Earth Syst. Dynam. 9 (2), 627–645.

Leavesley, G.H., Markstrom, S.L., Restrepo, P.J., Viger, R.J., 2002. A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling. Hydrol. Process. 16 (2), 173–187.

Lehner, B., Verdin, K., Jarvis, A., 2013. HydroSHEDS technical documentation version 1.2. EOS Trans. 89(10):26.

MacDonald, A.M., Bonsor, H.C., Dochartaigh, B.E.O., Taylor, R.G., 2012. Quantitative maps of groundwater resources in Africa. Environ. Res. Lett. 7 (24009).

MacDonald, A.M., Lark, R.M., Taylor, R.G., Abiye, T., Fallas, H., Favreau, G., Goni, I.B., Kebede, S., Scanlon, B., Sorensen, J.P.R., Tijani, M., Upton, K.A., West, C., 2021. Mapping groundwater recharge in Africa from ground observations and implications for water security. Environ. Res. Lett 16 (34012).

Manabe, S., 1969. Climate and the ocean circulation 1: I. The atmosphereic circulation and the hydrology of the Earth's surface. Mon. Weather Rev. 97 (11), 739–774.

Mao, J., Yan, B., 2019. Global monthly mean leaf area index climatology, 1981–2015. ORNL DAAC, Oak Ridge, Tennessee, USA Retrieved (https://daac.ornl.gov/VEGETATION/guides/Mean_Seasonal_LAI.html).

Moeck, C., Grech-Cumbo, N., Podgorski, J., Bretzler, A., Gurdak, J.J., Berg, M., Schirmer, M., 2020. A global-scale dataset of direct natural groundwater recharge rates: a review of variables, processes and relationships. Sci. Total Environ. 717 (137042).

Mohan, C., Western, A.W., Wei, Y., Saft, M., 2018. Predicting groundwater recharge for varying land cover and climate conditions-a global meta-study. Hydrol. Earth Syst. 22 (5), 2689–2703.

Niraula, R., Meixner, T., Ajami, H., Rodell, M., Gochis, D., Castro, C.L., 2017. Comparing potential recharge estimates from three land surface models across the Western US. J. Hydrol. 545, 410–423.

Oleson, K.W., Lawrence, D.M., Bonan, G.B., Drewniak, B., Huang, M., Koven, C.D., Levis, S., Li, F., Riley, W.J., Subin, Z.M., Swenson, S., Thornton, P.E., Bozbiyik, A., Fisher, R., Heald, C.L., Kluzek, E., Lamarque, J.F., Lawrence, P.J., Leung, L., Lipscomb, W., Muszala, S.P., Ricciuto, D.M., Sacks, W.J., Sun, Y., Tang, J., Yang, Z.L., 2013. Technical Description of Version 4.5 of the Community Land Model (CLM) 37.

Pelletier, J.D., Barron-Gafford, G.A., Breshears, D.D., Brooks, P.D., Chorover, J., Durcik, M., Harman, C.J., Huxman, T.E., Lohse, K.A., Lybrand, R., Meixner, T., McIntosh, J.C., Papuga, S.A., Rasmussen, C., Schaap, M., Swetnam, T.L., Troch, P.A., 2013. Coevolution of nonlinear trends in vegetation, soils, and topography with elevation and slope aspect: a case study in the Sky Islands of southern Arizona. J. Geophys. Res. Earth Surf. 118 (2), 741–758.

Pelletier, J.D., Broxton, P.D., Hazenberg, P., Zeng, X., Troch, P.A., Niu, G.Y., Williams, Z., Brunke, M.A., Gochis, D., 2016. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. J. Adv. Model. Earth Syst. 8 (1), 41–65.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100 (2), 81–92.

Quichimbo, E.A., Singer, M.B., Michaelides, K., Hobley, D., Rosolem, R., Cuthbert, M.O., 2021. DRYP 1.0: a parsimonious hydrological model of DRYland partitioning of the water balance. Geosci. Model Dev. 14 (11), 6893–6917.

Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., Döll, P., 2019. Challenges in developing a global gradient-based groundwater model (G3M v1.0) for the integration into a global hydrological model. Geosci. Model Dev. 12 (6), 2401–2418.

Reinecke, R., Wachholz, A., Mehl, S., Foglia, L., Niemann, C., Döll, P., 2020. Importance of spatial resolution in global groundwater modeling. Groundwater.

Reinecke, R., Müller Schmied, H., Trautmann, T., Andersen, L.S., Burek, P., Flörke, M., Gosling, S.M., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Thiery, W., Wada, Y., Yusuke, S., Döll, P., 2021. Uncertainty of simulated groundwater recharge at different global warming levels: a global-scale multi-model ensemble study. Hydrol. Earth Syst. Sci. 25 (2), 787–810.

Scanlon, B.R., Healy, R.W., Cook, P.G., 2002. Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol. J. 10, 18–39.

Siebert, S., Döll, P., Feick, S., Frenken, K., Hoogeveen, J., 2013. Global map of irrigation areas version 5 Rome.

Telteu, C.E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., Boulange, J.E.S., Andersen, L.S., Grillakis, M., Gosling, S.N., Satoh, Y., Rakovec, O., Stacke, T., Chang, J., Wanders, N., Shah, H.L., Trautmann, Mao, Hanasaki, N., Koutroulis, A., Pokhrel, Y., Samaniego, L., Wada, Y., Mishra, V., Liu, J., Döll, P., Zhao, F., Gädeke, A., Rabin, S.S., Herz, F., 2021. Understanding each Other's models an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. Geosci. Model Dev. 14 (6), 3843–3878.

Troch, P.A., Carrillo, G., Sivapalan, M., Wagener, T., Sawicz, K., 2013. Climate-vegetation-soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution. Hydrol. Earth Syst. Sci. 17 (6), 2209–2217.

Wada, Y., Van Beek, L.P.H., Van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P., 2010. Global depletion of groundwater resources. Geophys. Res. Lett. 37 (20), L20402.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5 (1), 13–26.

Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N.J.K., Pianosi, F., Rosolem, R., Stein, L., Woods, R.A., 2021. On Doing Hydrology with Dragons: Realizing the Value of Perceptual Models and Knowledge Accumulation. WIREs WATER e1550.

West, C., Rosolem, R., Macdonald, A.M., Cuthbert, M.O., Wagener, T., 2022. Understanding process controls on groundwater recharge variability across Africa through recharge landscapes. J. Hydrol. 612.

Williams, P.W., Ford, D.C., 2006. Global distribution of carbonate rocks. Zeitschrift Fur Geomorphologie, Supplementband 147 (1–2).

Zaherpour, J., Gosling, S.N., Mount, N., Müller Schmied, H., Veldkamp, T.I.E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., Wada, Y., 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environ. Res. Lett. 13 (6), 65015.