A first attempt to model global hydrology at hyper-resolution

Barry van Jaarsveld[1], Niko Wanders[1], Edwin H. Sutanudjaja[1], Jannis Hoch[1,2], Bram Droppers[1], Joren Janzing[3,4,5], Rens L.P.H. van Beek[1], and Marc F.P. Bierkens[1,6]

[1]Utrecht University, Department of Physical Geography, Princeton laan 8a, Utrecht, The Netherlands
[2]Fathom, Bristol, United Kingdom
[3]Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[4]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[5]Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, Switzerland
[6]Deltares, Unit Subsurface and Groundwater Systems, Utrecht, The Netherlands

Correspondence: Barry van Jaarsveld (a.s.vanjaarsveld@uu.nl)

Abstract. Global hydrological models are one of the key tools that can help meet the needs of stakeholders and policy makers when water management strategies and policies are developed. The primary objective of this paper is therefore to establish a first of its kind, truly global hyper-resolution hydrological model that spans a multiple-decade period (1985 - 2019). To achieve this, two key limitations are addressed, namely the lack of high resolution meteorological data and insufficient representation of lateral movement of snow and ice. Thus, a novel meteorological downscaling procedure that better incorporates fine-scale topographic climate drivers is incorporated, and a snow module capable of lateral movement of frozen water resembling glaciers, avalanches and wind movement is included. We compare this global 30 arc-seconds version of PCR-GLOBWB to previously published 5 arc-minutes and 30 arc-minutes versions by evaluating simulated river discharge, snow cover, soil moisture, land surface evaporation, and total water storage against observations. We show that hyper-resolution provides a more accurate simulation of river discharge, in particular for smaller catchments. We highlight that although global hyper-resolution modelling is possible with current computational resources and that hyper-resolution modelling results in more realistic representations of the hydrological cycle. However, our results also suggest that global hydrological modelling still needs to incorporate land cover heterogeneity and relevant hydrological processes at the sub kilometre scale to provide more accurate estimates of soil moisture and evaporation fluxes.

1 Introduction

Water is a vital and crosscutting element needed to achieve a number of sustainable development goals (Vörösmarty et al., 2015; Alcamo, 2019). Accurately simulating, predicting and forecasting the distribution, abundance, and shortage of water is therefore a crucial challenge for the hydrological community. By providing information on water resources, global hydrological models are one of the key tools that can help meet the needs of stakeholders and policy makers when water management strategies and policies are developed (Bierkens, 2015; Bierkens et al., 2015; Wood et al., 2011). Despite their usefulness, an ongoing critique is that the resolution of these models is unable to provide relevant information at scales at which adaptation strategies are implemented by stakeholders (Wada et al., 2017). In response to this criticism, there has been an effort to increase

the spatial resolution of current state-of-the-art global hydrological models and this push towards hyper-resolution hydrological models has previously been described as one of hydrology's "grand challenges" (Bierkens et al., 2015).

The drive to develop hyper-resolution models is based on the assumption that increased resolution will realise benefits coarse resolution counterparts can not. Hyper-resolution global hydrological models are expected to better capture the relevant physical processes that govern the distribution and quantity of global water resources and, therefore, provide a more detailed and accurate view of the hydrological cycle (Bierkens et al., 2015; Wood et al., 2011; Beven and Cloke, 2012). An improved and more detailed understanding of the global hydrological cycle can provide a number of important benefits to the broader scientific community and society. From scientific point of view, hyper-resolution hydrological models can facilitate progress and innovation in the fields of water quality, sediment transport, floods and drought risk by providing, much needed, detailed information on the movement of water in soils, rivers, lakes and ponds (Bierkens et al., 2015). Hyper-resolution hydrological models also promise benefits that will aid society as a whole, for instance high resolution hydrological data can provide stakeholders and policy makers with information on scales that are more logical and actionable (Bierkens et al., 2015; Wood et al., 2011; Beven et al., 2015).

To date, there has been noticeable progress towards a truly global hyper-resolution hydrological model; however, given the complexity of such an undertaking and the associated computational burden, hyper-resolution models have, so far, been confined to continental scale applications (e.g, Beven et al., 2015; Hoch et al., 2023; O'Neill et al., 2021; Vergopolan et al., 2021; Chaney et al., 2021). For example, the ParFlow model has been used to simulate groundwater and surface water for the contiguous United States at a spatial resolution of approximately 1 km (Yang et al., 2023). Also, for the contiguous United States, Aerts et al. (2022) analyzed how increasing the resolution from 3 km to 200 m in wflow_sbm affects predictions of river discharge. There has also been an attempt to model the European continent at the 1km resolution; Hoch et al. (2023) presents a 1 km version of PCR-GLOBWB which was used to simulate hydrological states and fluxes over a multi-decadal period. These studies have provided much needed headway towards truly global hyper-resolution modelling; but they have also brought to the fore a number of challenges that need to be overcome first.

Challenges surrounding global hyper-resolution models are related to epistemic uncertainties in input data and whether, or not, models at these finer resolutions can effectively capture and reproduce processes that govern water dynamics (Hoch et al., 2023; Aerts et al., 2022; Yang et al., 2023). Previous studies on continental scale hyper-resolution models have raised the question of whether an increased resolution actually results in a more accurate representation of the water cycle; there is mixed support for this notion. For example, when modelling at spatial resolutions approaching 1 km and comparing their accuracy to more coarse-scale counterparts, river discharge is more accurately simulated in some locations, while other locations show a reduced accuracy (Hoch et al., 2023; Aerts et al., 2022). Furthermore, there are discrepancies between how different components of the water cycle respond to an increase in the model resolution. Hoch et al. (2023) experienced that, as resolution increases, the fidelity of soil moisture and total evaporation as simulated with the global hydrological model decreases, even though river discharge shows an increase in accuracy when moving to finer resolutions (Hoch et al., 2023). How ever valuable the identification of such an inconsistency may be, perhaps more importantly, it provides for an opportunity to further understand how and why the different components of the water cycle respond to an increase in model resolution. Increases in model resolution have

also highlighted the need for the inclusion of fine-scale processes that are neglected at coarser resolutions. For example, Hoch et al. (2023) reports that in the absence of processes that represent the transport of frozen water through glaciers, avalanches and wind lead to unrealistic accumulations of frozen water accumulating in snow towers. Despite these challenges, continental scale hydrological models have shown that it is possible to accurately simulate at least some components of the hydrological cycle at resolutions approaching ≈1 km (Hoch et al., 2023; Yang et al., 2023), albeit not yet at the global scale.

One source of uncertainty is the mismatch between model resolution and that of meteorological data used as forcing (Hoch et al., 2023). The lack of meteorological data at the appropriate resolution is a major limitation for both coarse- and fine-scale models (Wilby et al., 2000; Benedict et al., 2019; Hoch et al., 2023; Yang et al., 2023; Döll et al., 2016; Müller Schmied et al., 2014). Available reanalysis products are created at a much coarser resolution than global hydrological models and fail to represent sub-grid climate dynamics that are important in defining local hydrological patterns. As a result, downscaling climate forcing becomes necessary for global hyper-resolution hydrological models and their accuracy is heavily dependent on how such downscaled products reflect reality. To date, the production of global climate models at resolutions discussed here are constrained by technical limitations around storing the large volumes of output data and computational resources required to complete such simulations (Schär et al., 2020; Karger et al., 2017). However, recently 1 km meteorological data have become available in the form of climatologies as in the case of WordClim (Fick and Hijmans, 2017) and CHELSA (Karger et al., 2017; Brun et al., 2022), which could feasibly be used to downscale coarse daily meteorological forcing data from reanalyses to the required hyper-resolution.

Given these considerations, there is a need to assess the feasibility of a truly global hyper-resolution hydrological model that relies on improved spatial representation of meteorological data and fine-scale hydrological processes. The primary objective of this paper is therefore to establish a first of its kind, truly global hyper-resolution hydrological model that spans a multiple-decade period (1985 – 2019); thereby extending the scope of current hyper-resolution hydrological models beyond the continental scale (Hoch et al., 2023). In this novel implementation of PCR-GLOBWB, a new downscaling procedure that better incorporates fine-scale topographic climate drivers are included. In addition, this implementation incorporates a snow module capable of lateral movement of frozen water resembling glaciers, which is pertinent at higher resolutions. We compare this global 30 arc-seconds simulation to previously published 5 arc-minutes and 30 arc-minutes versions of PCR-GLOBWB by evaluating simulated river discharge, snow cover, soil moisture, land surface evaporation, and groundwater storage against observation. We focus on how the model represents the hydrological cycle across scales and aim to highlight where we need to focus future efforts to improve hyper-resolution hydrological modelling.

2 Methods

2.1 30 arc-seconds PCR-GLOBWB Setup and Parametrization

PCR-GLOBWB (PCR - Global Water Balance) is a global hydrological and water resources model that estimates global water stores at various resolutions. It considers both natural and human-induced factors when estimating global water stores and fluxes. The 30 arc-seconds PCR-GLOBWB implementation presented here is built upon a schematization that has previously

been applied to continental Europe (Hoch et al., 2023), in such the model presented here largely follows that presented by Hoch et al. (2023), but with a significant increase in spatial expansion so that it now represents the entire globe.

The model parameterisation and inputs used in the 30 arc-seconds implementation represent high resolution hydrological processes where possible and in the following sections, we provide a summary of these. For extensive details on the setup of the 30 arc-seconds PCR-GLOBWB implementation, we refer the reader to the original European implementation by Hoch et al. (2023).

Land surface: soil, and cover, and topography

Soil information at the 30 arc-seconds resolution was derived from the SoilGrids250 dataset (Hengl et al., 2017), which is originally available at the 0.002 degree resolution. General soil attributes from SoilGrids250 were transformed into soil hydraulic properties, such as water holding capacity, field capacity, wilting point using the pedotransfer functions from Balland and Arp (2005). These properties were derived at the 0.002 arc-degree and upscaled to 30 arc-seconds, by averaging and using cell area as weights. For land cover parameterization the Global Land Cover Characteristics (GLCC) database version 2.0 (Loveland et al., 2000), with the land cover classification following Olson (1994a, b) and the parameter sets from Hagemann et al. (1999) and Hagemann (2002) were used. In addition the map of Global Food Security Support Analysis Data (GFSAD) version 1.0 (Teluguntla et al., 2016) was used to define irrigation areas at the 30 arc-seconds resolution. GLCC and GFSAD data are available at 30 arc-seconds resolution and thus one dominant land cover type was used for the 30 arc-seconds resolution. This is differs the 30 arc-minute and 5 arc-minute versions of PCR-GLOBWB, where each grid cell was divided into fractional constituents for four land cover types consisting of tall natural vegetation, short natural vegetation, non-paddy-irrigated crops, and paddy-irrigated crops (i.e. wet rice). The state-of-the-art Multi-Error-Removed Improved-Terrain Hydro digital elevation model (MERIT Hydro DEM; Yamazaki et al. (2019)) that is available at 3 arc-seconds resolution was used to derive topography related information. The 3 arc-seconds MERIT Hydro DEM was upscaled to the 30 arc-seconds, by averaging and using cell area as weights. It is important to note that, various sub-grid variability parameters, such as runoff-infiltration partitioning, interflow, groundwater recharge, and capillary rise, as well as evaporation processes (van Beek and Bierkens, 2008; van Beek, 2008; Hagemann and Gates, 2003; Todini, 1996) was derived at the 3 arc-seconds resolution and upscaled to the 30 arc-seconds, 5 arc-minutes and 30 arc-minutes resolution.

Surface water routing: lakes, reservoirs, and drainage/river network

Lakes and reservoirs information was taken from the Global Lakes and Wetlands Database (GLWD) of Lehner and Döll (2004) and the Global Reservoir and Dam Database (GRanD) of Lehner et al. (2011). The drainage network were adopted from the HydroSHEDS product (Lehner et al., 2008).

In brief, the model setup used here differs from the previous PCR-GLOBWB versions as follows: (i) the parallelisation approach used by the model is updated, (ii) a novel method of downscaling coarse-scale meteorological forcing to the required 30 arc-seconds resolution is incorporated, (iii) the model now allows for lateral transport of snow and ice at high elevations, and (iv) and an offline spin-up strategy is implemented. Together, these four changes to the model allowed us to complete a 30

arc-seconds PCR-GLOBWB simulation with a global extent by overcoming the computational hurdle whilst still maintaining enough similarity to the previously published versions so that model outputs can be compared and evaluated in a pragmatic way.

2.1.1 Climate Forcing Downscaling Procedure

Previously published 5 arc-minutes and 30 arc-seconds versions of PCR-GLOBWB relied on a lapse rate-centric approach to downscale meteorological forcing to the appropriate spatial resolution (Hoch et al., 2023; Sutanudjaja et al., 2011, 2018). In contrast, the current implementation relied on an alternative approach by making use of high resolution climatologies (Karger et al., 2017). The new downscaling methodology involved bilinearly interpolating the coarse-scale meteorological forcing data to the 30 arc-seconds resolution, followed by the calculation of monthly climatologies from the interpolated fields. Interpolated climatologies were then compared to monthly high resolution reference CHELSA climatologies (1981 - 2010; Karger et al. 2017 and Brun et al. 2022) to produce a set of Julian day-of-year correction factors that incorporated high resolution topographic information (Fig. 1). The high-resolution climatologies represent the years 1981 to 2010; as such, the correction factors were calculated for this time period.

Downscaling Temperature

As a first step, the coarse-scale daily temperature data (1981 - 2010) was interpolated to the 30 arc-seconds resolution using a bilinear interpolation (Tas_d). Thereafter, the interpolated values were used to calculate monthly climatologies (Tas_M) for the years 1981 - 2010 (Eq. 1); where N is the total number of years, M is the month and i is the day of month.

$$Tas_M = \frac{1}{N} \sum_{j=m}^{N} Tas_{d_{mi}} \tag{1}$$

The interpolated monthly climatologies were then compared to the high-resolution CHELSA reference climatologies ($Tas_{chelsa,M}$), using Equation 2, to obtain a set of monthly correction factors ($CF_{Tas,M}$).

$$CF_{Tas,M} = Tas_{chelsa,M} - Tas_M \tag{2}$$

Then, to obtain a correction factor for each Julian day of the year ($CF_{Tas,doy}$), where doy is day of year, we employed a linear interpolation on $CF_{Tas,M}$.

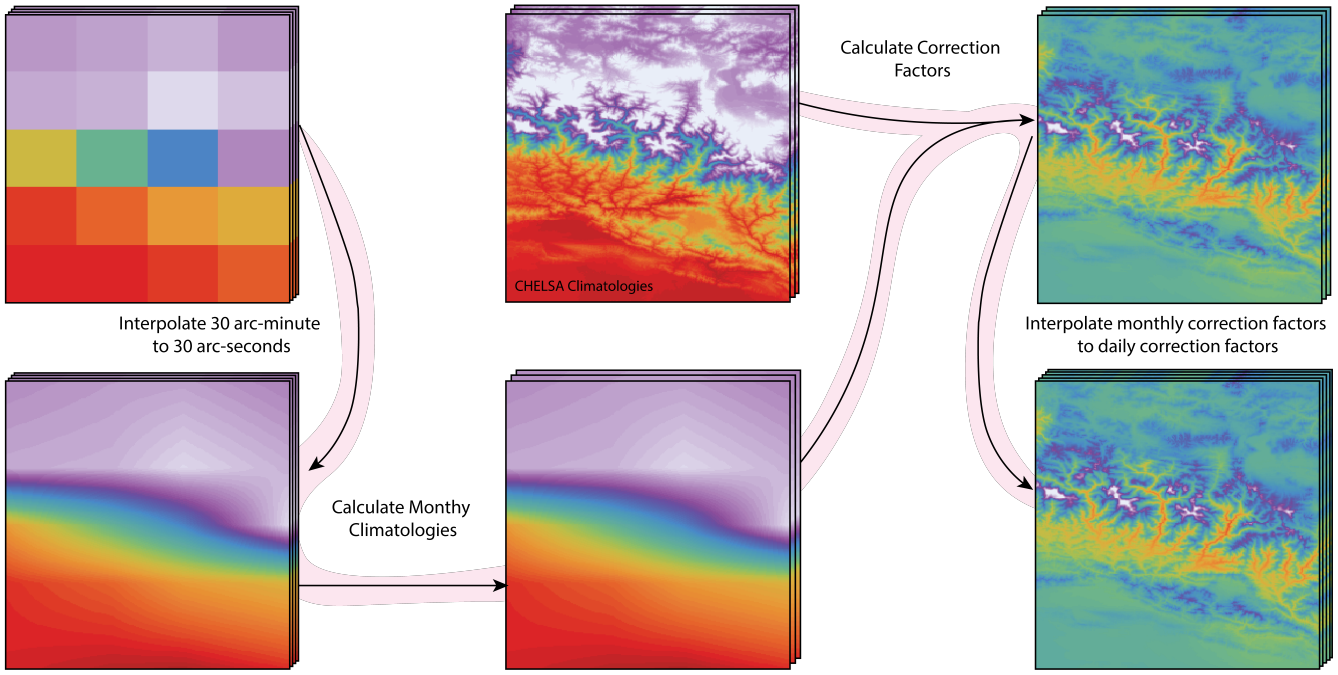

Figure 1. Procedure for downscaling meteorological forcing input date based on high resolution climatologies.

Downscaling Evaporation

Downscaling evaporation follows the same procedure as described above for temperature; where coarse scale daily data (1981 - 2010) was interpolated to the 30 arc-seconds resolution using a bilinear interpolation ($ET_{ref,d}$) and monthly climatologies ($ET_{ref,M}$) calculated for 1981 - 2010 (Eq. 3). Thereafter, interpolated monthly climatologies were then compared to the high resolution reference CHELSA climatologies ($ET_{chelsa,M}$) using equation 4 to obtain a set of monthly correction factors ($CF_{ET_{ref,M}}$). The final set of correction factor for each Julian day of the year ($CF_{ET,doy}$) was obtained through the linear interpolation of $CF_{ET_{refM}}$. The use of a multiplicative correction factor here was in order to handle variance conservation and to ascertain strictly positive values.

$$ET_{ref,M} = \frac{1}{N}\sum_{j=m}^{N} ET_{ref,d_{mi}} \tag{3}$$

$$CF_{ET_{ref,M}} = ET_{chelsa,M}/ET_{ref,M} \tag{4}$$

Downscaling Precipitation

As a first step, coarse-scale daily data (1981 - 2010) was interpolated to the 30 arc-seconds resolution using a bilinear interpolation (Tp_d). For the precipitation downscaling, an additional step was necessary to correct for drizzle days. Drizzle days are erroneous by-products from interpolating precipitation, which result in very light precipitation where precipitation should be zero. To account for this and remove excess precipitation, we calculated which proportion of days in each month of the year are dry days ($dryDays$) and set that proportion of bottom values in the interpolated precipitation product to 0 (Eq. 5).

$$Tp_d = \begin{cases} Tp_d & \text{if } Tp_d \text{ percentile rank is > than } dryDays \\ 0 & \text{if } Tp_d \text{ percentile rank is < than } dryDays \end{cases} \tag{5}$$

Thereafter, the interpolated values were used to calculate climatologies (Tp_M) from 1981 - 2010 (Eq. 6). The interpolated monthly climatologies were then compared to the high resolution reference climatologies ($Tp_{chelsa,M}$) using equation 7 to obtain a set of monthly correction factors (CF_{Tp_M}).

$$Tp_M = \frac{1}{N}\sum_{j=m}^{N} Tp_{d_{mi}} \tag{6}$$

$$CF_{Tp_M} = Tp_{chelsa,M}/Tp_M \tag{7}$$

Then, to get a correction factor for each Julian day of the year ($CF_{Tp,doy}$) we employed a linear interpolation on CF_{Tp_M}. Again, as with evaporation, the use of a multiplicative correction factor here was in order to handle variance conservation and to ensure that precipitation is positive.

2.1.2 Snow and Ice transport to mimic glaciers, avalanches and wind transport

A limitation of PCR-GLOBWB highlighted by Hoch et al. (2023) relates to how the model handles snow and ice at high elevations. Downscaled temperatures are rarely above freezing point at higher elevations, and given that snowmelt is calculated using the degree day model, this results in unrealistic accumulations of frozen water. In reality excess snow and ice would be transported downslope by glaciers, avalanches and wind; however, these processes are not captured in the previous versions of PCR-GLOBWB. To solve this, we included a mechanism that allows lateral movement of frozen water to mimic the lateral and downslope transport of snow and ice by glaciers, avalanches and wind. The snow and ice distribution component implemented here largely follows that described by Frey and Holzmann (2015). If the snow water equivalent exceeds a threshold of (Hv = 0.625 m), lateral transport is activated. When transport is activated, the excess (i.e. transportable) volume of frozen water in a donor cell, $snow_d$, is calculated from Equation 8 and is then distributed to neighbouring down slope acceptor cells as a function of slope steepness (Eq. 9). A similar approach has previously been implemented in the community water model (Burek et al., 2020).

$$snow_D = max(SWE - Hv, 0) * cellArea \tag{8}$$

$$snow_A = \frac{snow_D * \frac{tan(slope)}{90}}{N_{acceptorCells}} \tag{9}$$

2.1.3 Spin up strategy

Traditionally, to get an initial estimate of the water storage and fluxes, PCR-GLOBWB requires a mandatory spin-up period, during which the model is simulated for the first time step repetitively until the hydrological storage values (e.g., unsaturated and saturated zone) have converged to long-term steady states. However, when considering the computational resources required for a global 30 arc-seconds simulation, this approach becomes unfeasible, as times to reach equilibrium values would be very large, which is especially true for states that evolve slowly (i.e., groundwater storage). To overcome this obstacle, a three-phase spin-up process is implemented in the 30 arc-seconds schematization. In the first phase, PCR-GLOBWB is run for a three-year period to obtain a representative annual groundwater recharge rate (Gw_{rech}). Groundwater storage is then calculated in the same way as is done by the complete model using Equation 10 for 1 000 iterations, starting with a values of $1x - 10$ where the base flow is driven by the response time of the groundwater aquifer (j). In the third and final phase, the model is run for an additional period of at least 5 years with the precalculated groundwater storage values as initial conditions to obtain the final set of initial conditions.

$$GW_{stor,i} = GW_{stor,i-1} + GW_{rech_i} - [j \times \overline{GW_{stor}} \times (\frac{GW_{stor,i-1}}{\overline{GW_{stor}}})^1] \tag{10}$$

2.1.4 Parallelisation Approach

Maintaining pragmatic and feasible simulation times is a significant challenge when considering hyper-resolution simulations. A simple yet effective parallelisation technique used in previous PCR-GLOBWB implementation is to spatially partition the modelling domain into independent hydrological units and assign separate processors to each unit which are then completed concurrently. In the previous 5 arc-minutes PCR-GLOBWB, 53 independent spatial hydrological units were completed in parallel (Fig. 2a). This is possible because each basin's outlet ends up in a reservoir, endorheic lake, or ocean. This approach was followed in the current 30 arc-seconds implementation, where the modelling domain is split up into 215 independent hydrological units, which can be completed in parallel (Fig 2b). However, at 30 arc-seconds, for some of the larger basins in the domain, this approach still leads to extremely long simulation times if not subdivided further - predominantly because of computationally expensive calculations associated with surface water routing. For basins exceeding an 800 000 km^2 threshold, a hierarchical method of parallelisation was therefor used. This threshold was selected to balance efficient input/output operations and the number of point operations done by an individual processor. First, the basin is divided according to stream order so that each subbasin is smaller than the 800 000 km^2 threshold. The upper reaches of the basin are completed first and then followed by the next downstream subbasin until the last subbasin has all the necessary information (Fig 2c).

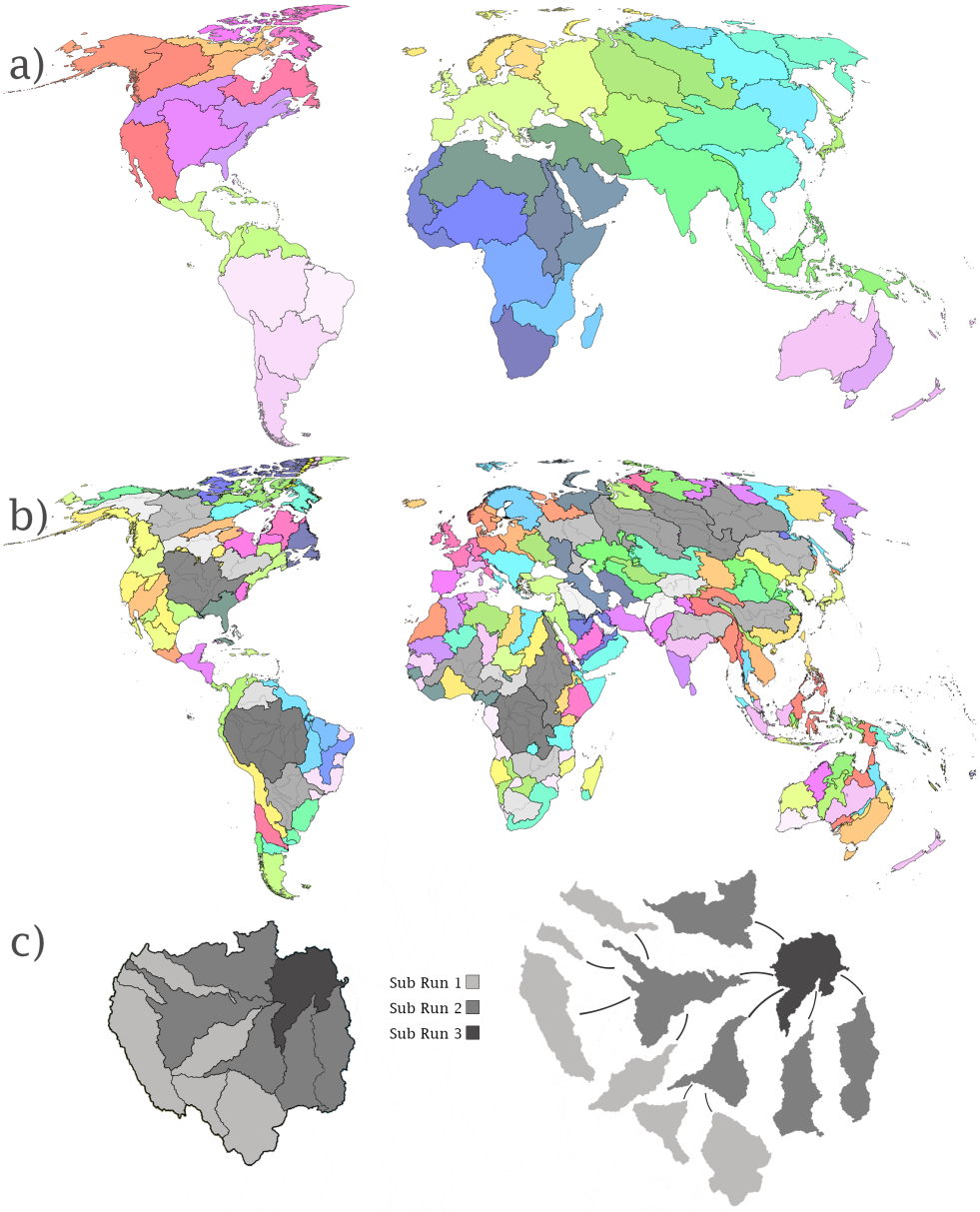

Figure 2. (a) Spatial domains used for parallelisation in the previously published 5 arc-minutes PCR-GLOBWB version and (b) the domains used for the parallelisation approach used in this 30 arc-seconds PCR-GLOBWB; which (c) requires additional dividing and hierarchical parallelisation for basins exceeding 800 000 km^2 (which are displayed in grey in (b)). Using the Amazon as an example, the hierarchical parallelisation involves simulating the upper sub-basins (Sub Run 1) first, followed by the intermediate sub-basins (Sub Run 2), and finally the penultimate sub-basin (Sub Run 3).

2.2 Global 30 arc-seconds PCR-GLOBWB simulation and evaluation

2.2.1 Simulation

The global 30 arc-seconds parameterization described above was simulated for a multi-decadel period (1985-2019) and forced with downscaled 30 arc-minutes W5E5 temperature, precipitation, and reference potential evaporation (Lange et al., 2021). The reference potential evaporation was calculated from the Penman-Monteith formulation, following the FAO guidelines (Allen and Food and Agriculture Organization of the United Nations, 1998), using the python package pyEt (Vremec and Collenteur, 2021); input data for the calculation of the reference potential evaporation was also taken from W5E5. The initial conditions for this simulation were calculated following the three part spin-up approach described above. As the first phase, PCR-GLOBWB was run from 1979 - 1981 with hydrological states set at an initial of 0.001 m of water and 1981 was taken as the representative groundwater recharge year to calculate groundwater storage offline. The model was then put through an additional spinup period of 6 years (1979 - 1985) to get stable estimates of the other fluxes and storages. For final production, the initial conditions were used to run PCR-GLOBWB from 1985 - 2019. All simulations in this paper were run on Snellius, the Dutch National supercomputer. We note that no calibration was performed for any of the simulations reported in this study.

To assess how capable this 30 arc-seconds PCR-GLOBWB is at reproducing the global hydrological cycle compared to coarser versions, two additional simulations were performed; one at the 5 arc-minutes and the other at the 30 arc-minutes. The same forcing and, where applicable, the same model settings were used. A spinup period of 40 years was used. For a more thorough explanation of these two models, we refer the reader to the original publications for the 30 arc-minutes (van Beek, 2008; van Beek and Bierkens, 2008; Van Beek et al., 2011) and 5 arc-minutes variants of PCR-GLOBWB (Sutanudjaja et al., 2018). All three models were simulated for the same time period, 1985 - 2019, and used the same meteorological forcing; however, key differences between the models versions are highlighted in Table 1.

Table 1. Table of key differences between the 30 arc-minutes, 5 arc-minutes and 30 arc-seconds PCR-GLOBWB implementations.

	30 arc-minutes	5 arc-minutes	30 arc-seconds
Downscaling Procedure	No	Lapse-Rate	Climatologies
Land cover sub grid variability	Yes	Yes	No
Lateral Snow Transport	No	No	Yes
Parallelisation	No	Basin Level	Basin and sub-basin level

2.2.2 Evaluation

To provide a more comprehensive evaluation of the global simulations, multiple hydrological variables were used for evaluation, namely total water storage, total evaporation, soil moisture, snow cover, and river discharge. To compare simulations of different resolution to one another and observation data in a fair way, we opted to evaluate the simulations at scales that matched those of the observational data. In doing so, it allowed for an assessment of the simulated values, regardless of simu-

lation resolution, whilst still allowing for comparisons between scales.

Global Water Balance and Total Water Storage

As a first comparisons between the different resolutions, we also calculated the global water balance and its respective components. In order to determine to what degree the models are able to partition water into different components of the water cycle, the global water budgets (Eq. 11) were calculated for 1985 - 2019.

$$\mathrm{P} = \mathrm{E} + \mathrm{Q} + \Delta\mathrm{S} \tag{11}$$

where, P: Precipitation, E: Evaporation, R: Runoff, ΔS: delta storage.

This allowed us to compare mean annual fluxes of precipitation, evaporation, runoff and change in storage for the different resolutions. In addition, runoff/precipitation and evaporation/precipitation ratios were calculated for simulations at the three different resolutions.

Monthly simulated total water storage was evaluated against JPL TELLUS GRACE/GRACE-FO data for 2002 - 2019 (Kornfeld et al., 2019). Given the large difference in spatial resolution between the GRACE data and the simulated data, this evaluation was conducted at the basin level. GRACE data used here are at the 30 arc-minutes resolution, which is close to the Mascon solution provided in Kornfeld et al. (2019), however the original resolution of GRACE is 3 arc-degrees. Therefore we aimed to produce a global map of basins so that each basin contains at least 4 grid cells at the original 3 arc-degree (i.e., the footprint of the original 3 arc-degree GRACE observations). Basin outlines were obtained from HydroBasins (Lehner and Grill, 2013), aggregation level 3, and all basins smaller than 400 000 km^2 were merged to neighbour basins exceeding 400 000 km^2. Basins that could not be merged, such as small islands, were removed. For each basin, the relative root mean sum of squares (RRMSE; Eq. 12 and Spearman's rank correlation coefficient; Eq. 13) were calculated, as an indication of how well the model was able to reproduce the temporal patterns and magnitude of total water storage anomalies respectively.

$$RRMSE = \frac{RMSE(obs, sim)}{\sigma(obs)} \tag{12}$$

where $RMSE$: Root mean square error between observations (obs) and simulations (sim) and $\sigma(obs)$: is the standard deviation of the observations.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{13}$$

where ρ: Spearman's rank correlation coefficient, d_i: difference between the two ranks of each observation, and n: number of observations.

Total Evaporation and Soil Moisture

Total evaporation and soil moisture were evaluated against station based observation data. Soil moisture data was obtained from the International Soil Moisture Network (Dorigo et al., 2021). In cases where multiple soil moisture measurements were present within a single day, the daily mean was calculated and used for evaluation. In addition, to ensure a good match between the modelled soil moisture depths and the observations, only data that coincided with the depth of the first soil layer were used. To match the location of observation stations with the appropriate grid cells, we located the nearest grid cell relative to the coordinates of the observation station.Observed total evaporation was obtained from the FLUXNET data set (Pastorello et al., 2020). As with soil moisture, the observed values were matched to the simulated data by locating the nearest grid cell.

The Kling-Gupta Efficiency (KGE; Eq. 14) was used to assess the accuracy of the simulated variables and for both total evaporation and soil moisture, evaluation was restricted to stations with at least 1 095 days of observation data. KGE values range from -∞ to 1.0, with values greater -0.41 implying that the model is a better predictor than the mean of the data (Knoben et al., 2019).

$$KGE = 1 - \sqrt{(\rho - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{14}$$

In addition, we analysed how the different components of the KGE score differed between resolutions. Correlation coefficients (ρ) provide an overview of how well the model reproduces temporal changes in the observed data, bias ratio (β) indicates differences between the means of the simulated and observed values, and variability ratio (α) indicates how well the model replicates the variability of the observed data. A perfect KGE score is 1, which arises when all components of the score ρ, α and β equal 1 (i.e, the observed and modelled values are identical). It is important to note that for both observed evaporation and soil moisture are not uniformly represented across the modelling domain. Observations are denser over North America and the European continent compared to the rest of the world (Fig A1 & Fig. A2).

Snow Cover

To establish to what degree the simulations were able to reproduce snow dynamics, we evaluated daily snow cover at the 30 arc-seconds resolution, using the MODIS daily snow cover product as observation data (Nagler et al., 2021). For this, simulated snow water equivalent was converted into snow cover, where values greater than zeros were classified as having snow present and assigned a value of one, while values equal to zero were classified as having no snow and assigned a value of zero. Given the mismatch in spatial resolution between the observation data (30 arc-seconds) and the 5 and 30 arc-minutes simulations, the 5 arc-minutes and 30 arc-minutes simulated snow cover were re-gridded to the 30 arc-seconds resolution using the nearest-neighbour algorithm. As an estimate of accuracy, we calculated the false alarm rate (FAR; Eq. 15), probability of detection (POD; Eq. 16), success ratio (SR; Eq. 17) and brier score (Eq. 18). POD (perfect score = 1) indicates the probability of the observed snow events to be correctly forecasted; whereas, FAR (perfect score = 0) indicates which fractions of the simulated snow events incorrectly simulated the presence of snow when there was no snow in observed data. The SR (perfect score = 1) gives information on the fraction of the observed snow events were correctly forecasted. The brier score (perfect score = 0) was calculated for an overall assessment of the magnitude of error. These scores could conceivably be calculated for each

grid cell over the entire domain; however this would then include large regions that never experience snow and ultimately skew validation scores. Therefore, we limited the validation to regions that actually experience snowfall in reality. To do this objectively, we created a mask where snowfall was present in the simulation or the MODIS observation data and constrained the validation within the mask.

$$\text{FAR} = \frac{\text{false alarms}}{\text{correct negatives} + \text{false alarms}} \tag{15}$$

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}} \tag{16}$$

$$\text{success ratio} = \frac{\text{hits}}{\text{hits} + \text{false alarms}} \tag{17}$$

$$\text{brier score} = \frac{1}{N}\sum_{t=1}^{N}(sim_t - obs_t)^2 \tag{18}$$

where sim_t: is the presence or absence of snow cover in the simulation, obs_t: is the presence or absence of snow cover in the observed data for time step t.

River Discharge

River discharge from the Global Runoff Data Center (GRDC) was used to evaluate river discharge simulated by PCR-GLOBWB for the three resolutions. The stations used for evaluation met the following criteria (i) 1 095 days of daily data were available and (ii) the catchment area was greater than 5 km^2. Also, to make sure that the observations locations were coupled to the right tributary, we selected the relevant grid cell by matching the catchment area reported in GRDC with that of the catchment area used in the model. The grid cell within a 5 km window of the station coordinate which had a catchment area closest to that reported by GRDC was selected as the representative point. To enable comparisons between different resolutions, stations were selected based on the catchment area of the of 30 arc-seconds modelling domain. As the evaluation statistic, we calculated the KGE (Eq. 14) and used the value of -0.41 as the boundary value to determine whether the model improves upon the mean variable benchmark. A perfect KGE score is 1.0, and values greater that -0.41 indicate that the model is a better predictor that using the variable mean value (Knoben et al., 2019).

In addition, to obtain information on how the 30 arc-seconds simulation compared in relation to the 5 arc-minute and 30 arc-minute counterparts, we calculated the KGE skill score (Eq. 19). This allows for inference on whether a simulation improved compared to a benchmark simulation (Towner et al., 2019). Here we assessed how the 30 arc-seconds simulations potentially improved upon the 30 arc-minutes and 5 arc-minutes simulations as benchmarks; where a positive value indicates

an improvement, and a negative value indicates a regression. To facilitate visualisation and a more intuitive comparison, the bounded variant of the KGE was used so that evaluation scores range from -1 to 1 (Hallouin, 2021). In addition, this bounded variant was used to explore its relationship between elevation and catchment area; elevation was obtained from the digital elevation model within PCR-GLOBWB, and catchment area as reported in the GRDC database.

$$KGE_{ss} = \frac{KGE_a - KGE_{ref}}{1 - KGE_{ref}} \tag{19}$$

3 Results

3.1 Increased resolution: computational load and insights

Increased model resolution is associated with significant increases in computational load and storage requirements (Table 2). Without parallelisation a global 5 arc-minutes and 30 arc-seconds would be impractical, with simulation times taking months. However, with the inclusion of a previously implemented basin level (Sutanudjaja et al., 2018) and the sub-basin level parallelisation scheme developed in this study, simulation times are now in the orders of days ($\approx$ 2.5 days) for the 5 arc-minutes resolution and weeks ($\approx$ 17 days) for the 30 arc-seconds resolution. We also find that, unsurprisingly, storage requirements increases rapidly with increased spatial resolution. For instance, writing a single variable to disk at the daily frequency results in a 2 000 fold increase in storage requirements when moving from 30 arc-minutes to 30 arc-seconds (Table 2). Storing all possible variables will equate to 365 Gb for the 30 arc-minutes while the 30 arc-seconds resolution storage requirements approach the petabyte scale.

Table 2. Overview on the computational and storage requirements for a multi-decadel (1985 - 2019) global PCR-GLOBWB simulation at different resolutions 30 arc-minutes, 5 arc-minutes and 30 arc-seconds resolutions. Simulations were run on the Dutch National Supercomputer - Snellius.

Resolution	Serial Simulation Time (hrs)	Parallel Simulation Time (hrs)	Storage (GB; min - max)	max CPU cores
30 arc-minute	43	not applicable	3.3 - 365	1
5 arc-minute	2 465	66	182 - 20 748	1 696
30 arc-seconds	39 245	401	8 200 - 934 800	6 528

Increased model resolution does, however, provide a unique and improved representation of the hydrological cycle (Fig. 3). For instance, compared to the 5 arc-minutes and 30 arc-seconds resolution, the 30 arc-seconds resolution better resolves spatial patterns in soil saturation related to elevation and land cover type. In the 30 arc-seconds simulation, variations in soil saturation degree related to drainage networks become visible. When comparing the resolutions over the Himalayas, the 30 arc-second simulation represents a continuous field and does not display the footprint of the original coarse scale meteorological forcing as is present in the 5 arc-minute and 30 arc-minute resolutions (Fig. 3). In addition, when shifting focus towards the Taihang

mountains in mainland China, the difference between the arid Gobi towards the west and humid forests towards the east is most evident in the 30 arc-second simulation (Fig. 3).

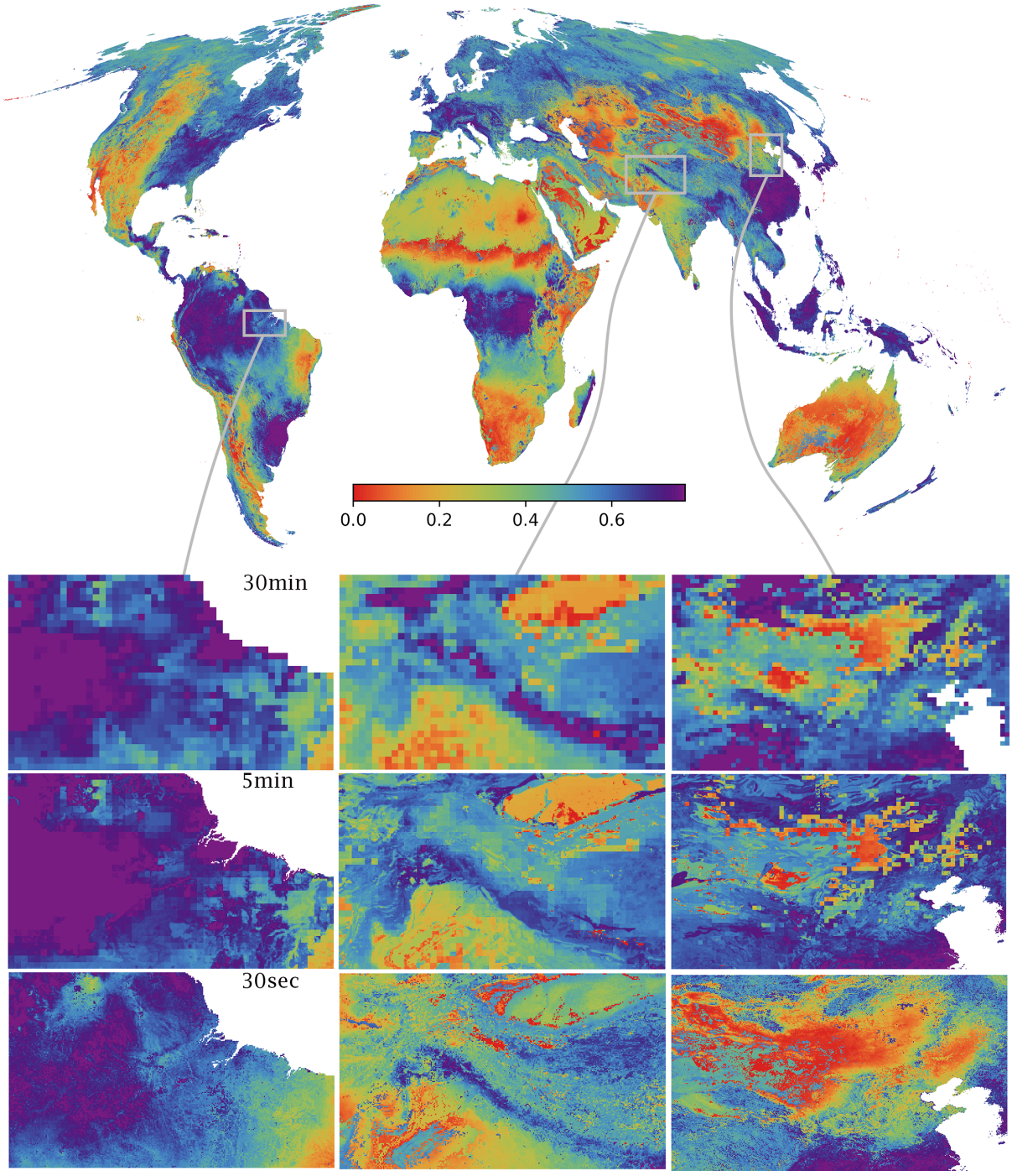

Figure 3. Mean (1985 - 2019) upper soil saturation (-) simulated by PCR-GLOBWB at the 30 arc-seconds resolution. Zoomed insets show how the highlighted regions differ between the 30 arc-seconds, 5 arc-minutes and 30 arc-minutes resolution PCR-GLOBWB simulations.

3.2 Global Water Balance and Total water storage

To evaluate the impact of increasing the spatial resolution on the representation of the global water cycle, we compared the long-term averages in the global water balance components. In terms of absolute values, the 30 arc-minute and 5 arc minutes resolutions have comparable long term averages for the components of the water balance (Table 3); differences are attributable to variation in the total land area being simulated. In contrast, the 30 arc-seconds resolution shows higher precipitation and evaporation and lower runoff (Table 3).

To allow for easier comparison, we focus on relative differences. For the 30 arc-minutes and 5 arc-minutes the results are comparable; however, when considering the 30 arc-seconds simulation, the relative evaporation rates are significantly higher and runoff significantly lower compared to the other two simulations (Table 3). As resolution increases, there is a decrease in the relative amount of precipitated water that accumulates as runoff, which is due to the increase in the relative rates of evaporation (Fig. 4).

Table 3. Global water balance components $km^3 year^{-1}$ for 1985 - 2019 for a 30 arc-minutes, 5 arc-minutes, and 30 arc-seconds simulated by PCR-GLOBWB. Values in parenthesis indicate the relative amounts in relation to precipitation as a percentage.

Simulation	Precipitation	Evaporation	Runoff	Storage
30 arc-minutes	112 140	64 091 (57%)	47 089 (42%)	960
5 arc-minutes	112 159	63 277 (56%)	49 223 (44%)	-341
30 arc-seconds	121 798	84 275 (69%)	39 605 (33%)	-2 082

Spatial patterns in the models ability to reproduce total water storage anomalies are similar in the fact that for all three resolutions continental arid and high elevation regions are less well represented in comparison to low laying temperate, more mesic continental regions and islands (Fig. 5a). Overall, we find that, in terms of temporal dynamics, the 5 arc-minutes resolution best resembles the GRACE data, followed by the 30 arc-seconds and 30 arc-minutes resolution (Fig. 5b). Regarding the magnitudes of errors, the 5 arc-minutes resolution best resembles the GRACE data, followed by the 30 arc-minutes and 30 arc-seconds resolution (Fig. 5c).

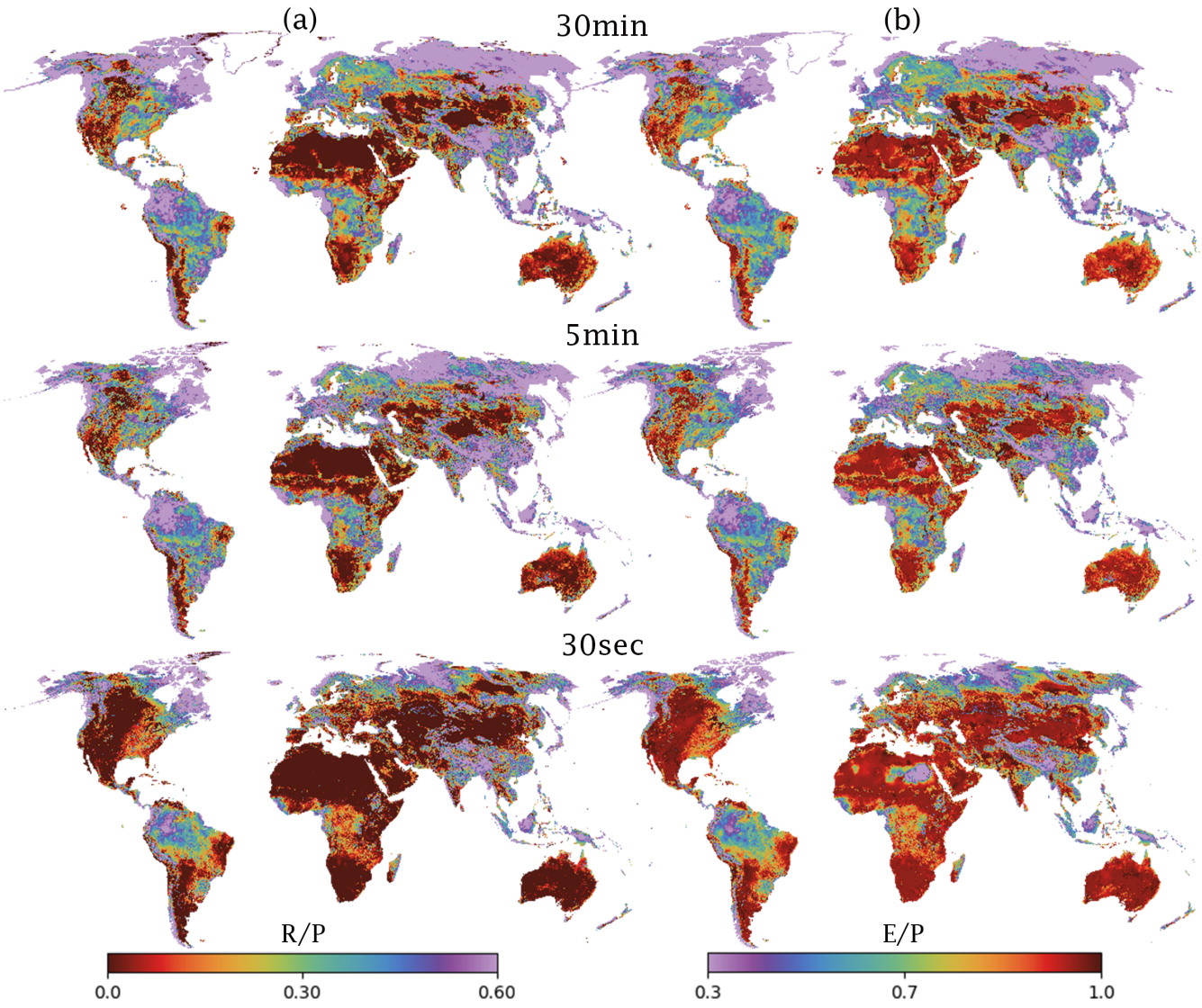

Figure 4. Mean annual mean (a) Runoff/Precipitation and (b) Evaporation/Precipitation ratios for a 30 arc-minutes, 5 arc-minutes, and 30 arc-seconds simulated by PCR-GLOBWB from 1985 - 2019.

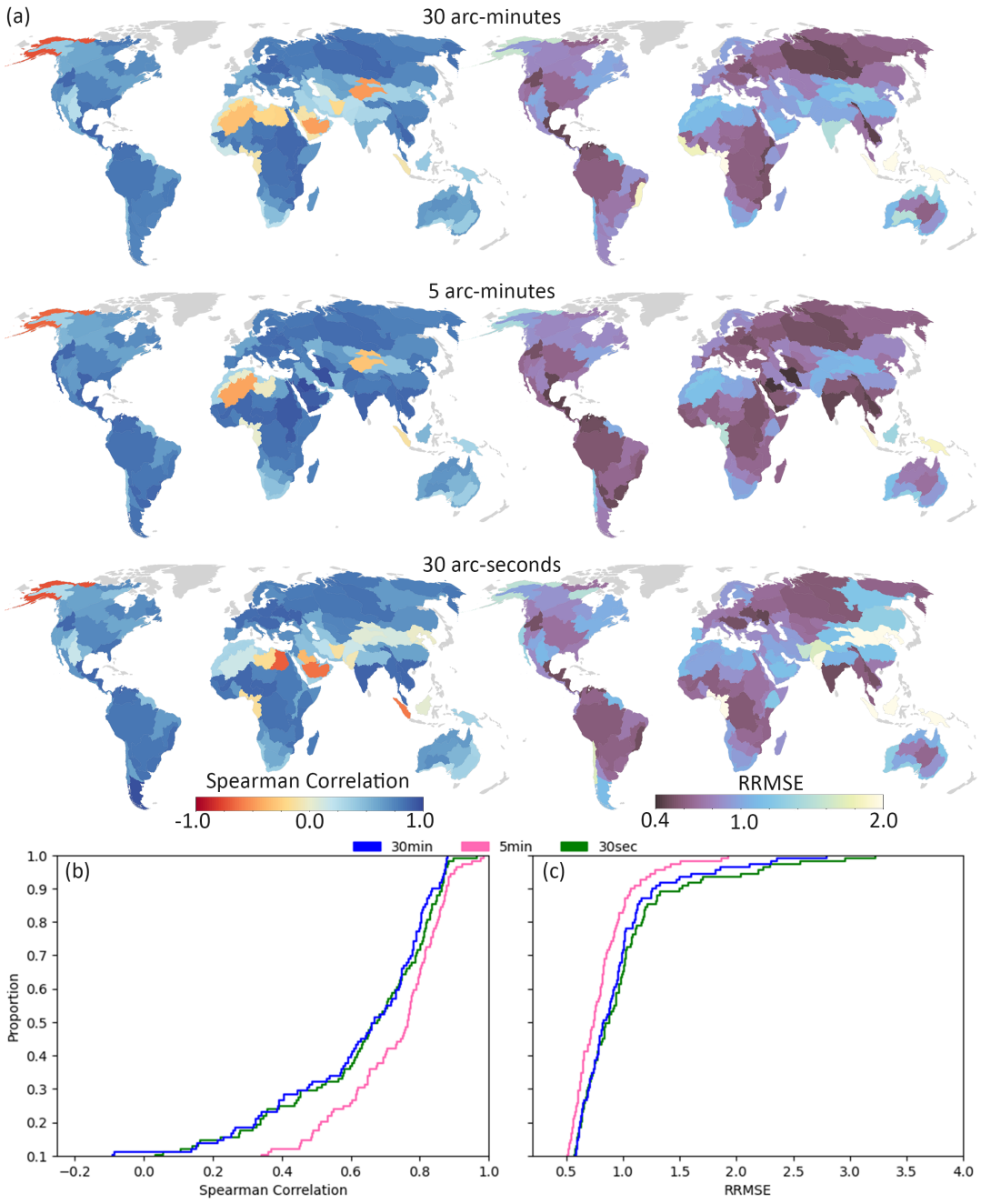

Figure 5. (a) Spatial variation of of spearman correlation and RRMSE for simulated total Water Storage validated against GRACE data, from 2002 - 2019, for basin which areas exceed 400 000 km^2. Empirical distribution functions for (b) spearman correlatiosn and (c) RRMSE.

3.3 Soil Moisture and Total Evaporation

The cumulative frequency curves of KGE of simulated soil moisture are largely indistinguishable between the 30 arc-minutes, 5 arc-minutes and 30 arc-seconds resolutions. Of the 1 676 stations used for validation (Fig. A1), 50% display a KGE greater than -0.41 (Fig. 6a) - this is true for all three simulations. Yet, when considering the different components of KGE, differences between model resolutions are evident. As resolution increases the correlation increases (Fig. 6b) and variability decreases (Fig. 6d); yet this is offset by a increase in bias (Fig. 6c). In contrast, total evaporation displays significant differences between the 30 arc-seconds simulation in comparison to the 5 arc-minutes and 30 arc-minutes simulation; whereas evaporation for the 30 arc-minutes and 5 arc-minutes are similar. Only 85% of the 143 stations (Fig. A2) display a KGE greater than -0.41 for the 30 arc-seconds resolution; whereas, almost all of the stations in the 30 and 5 arc-minutes simulation exceeded this threshold (Fig. 6e). This difference is attributable to an overestimation of the mean bias (Fig. 6g) and variance (Fig. 6h) of the simulated evaporation compared to observations.

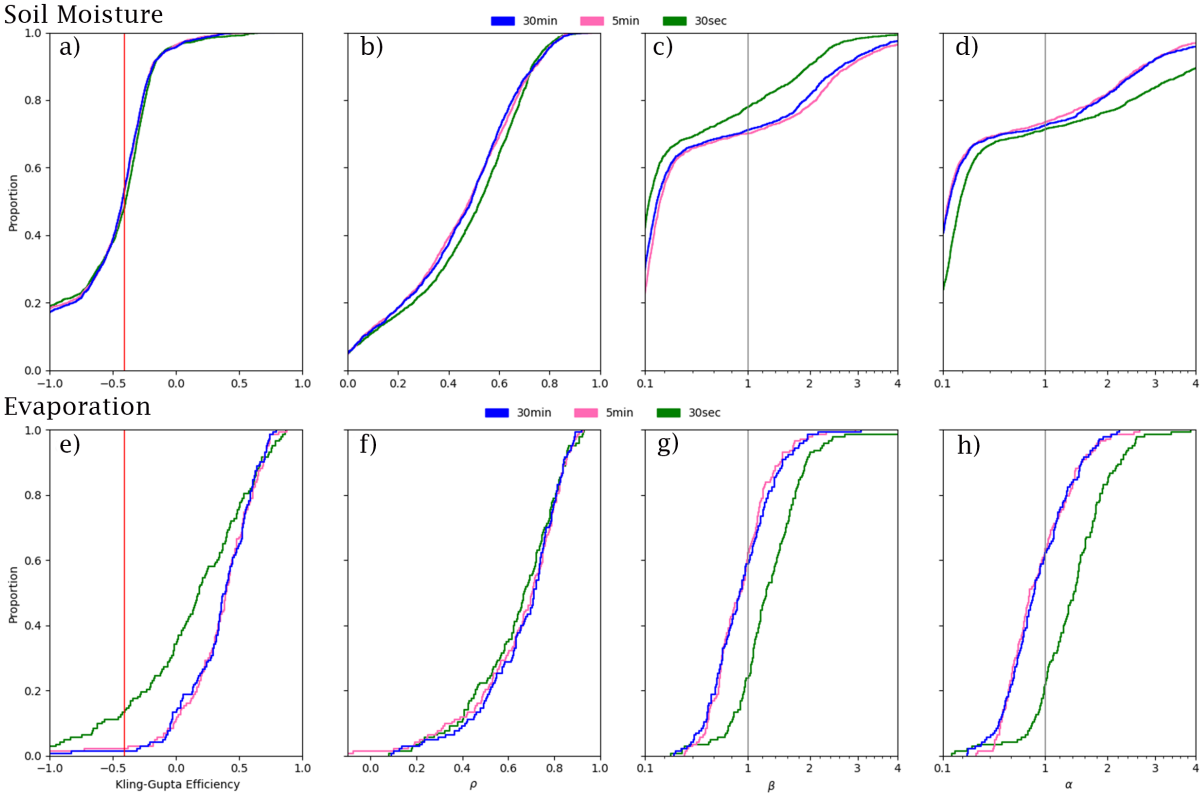

Figure 6. Top: Soil Moisture (a) KGE and associated constituents of KGE: (b) correlation, (c) beta, (d) alpha from the ISMN. **Bottom:** Total Evaporation (e) KGE and associated constituents of KGE: (f) correlation, (g) beta, (h) alpha from FLUXNET. Values greater than -0.41 (red line) indicate the value at which stations improves upon the mean flow benchmark (Knoben et al., 2019).

3.4 Snow Cover

The inclusion of the lateral movement of frozen water, representative of movement by glaciers, avalanches or wind, resulted in a more accurate accumulation and redistribution of frozen water at high elevation and prevented the erroneous accumulation of frozen water into snow towers. When evaluated against observed snow cover, all three the resolutions tended to overestimate the frequency of snow cover. The 30 arc-minutes simulation displays the highest false alarm rate and shows an increased tendency to simulate snow on occasions where there is no snow present (Fig 7a). As resolution increases, the false alarm rate decreases, so that the simulations of 30 arc-seconds display the lowest false alarm rate (Fig 7a). The success rate follows the inverse pattern of the false alarm rate, where increasing resolution results in decreased success rates (Fig 7b). With a highest probability of detection, the 5 arc-minutes simulation shows the best ability to correctly simulate occasions where snow is present, followed by the 30 arc-seconds and 30 arc-minutes simulation (Fig 7c). As an overall assessment of the resolutions to correctly simulate the presence and absence of snow, the brier score reveals that the 30 arc-second resolutions are the most accurate, followed by the 5 arc-minutes and 30 arc-minutes resolutions (Fig. 7 d).

The global distributions in the brier score reveals the performance varies substantially depending on location (Fig. 8). For all three resolutions, the models show the poorest performance in arid regions and regions of high elevation and complex topography. In terms of differences between resolutions, the performance increase with resolution is highest for higher elevations. However, for drier regions the performance degrades (Fig. 8).

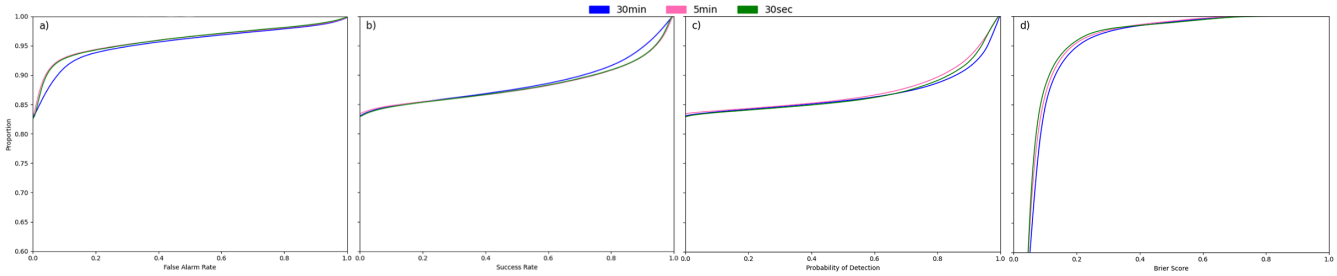

Figure 7. (a) False alarm rate, (b) success rate, (c) probability of detection and (d) brier score for daily snow cover for a 30 arc-minutes, 5 arc-minutes and 30 arc-seconds global PCR-GLOBWB simulation from 1985 - 2019.

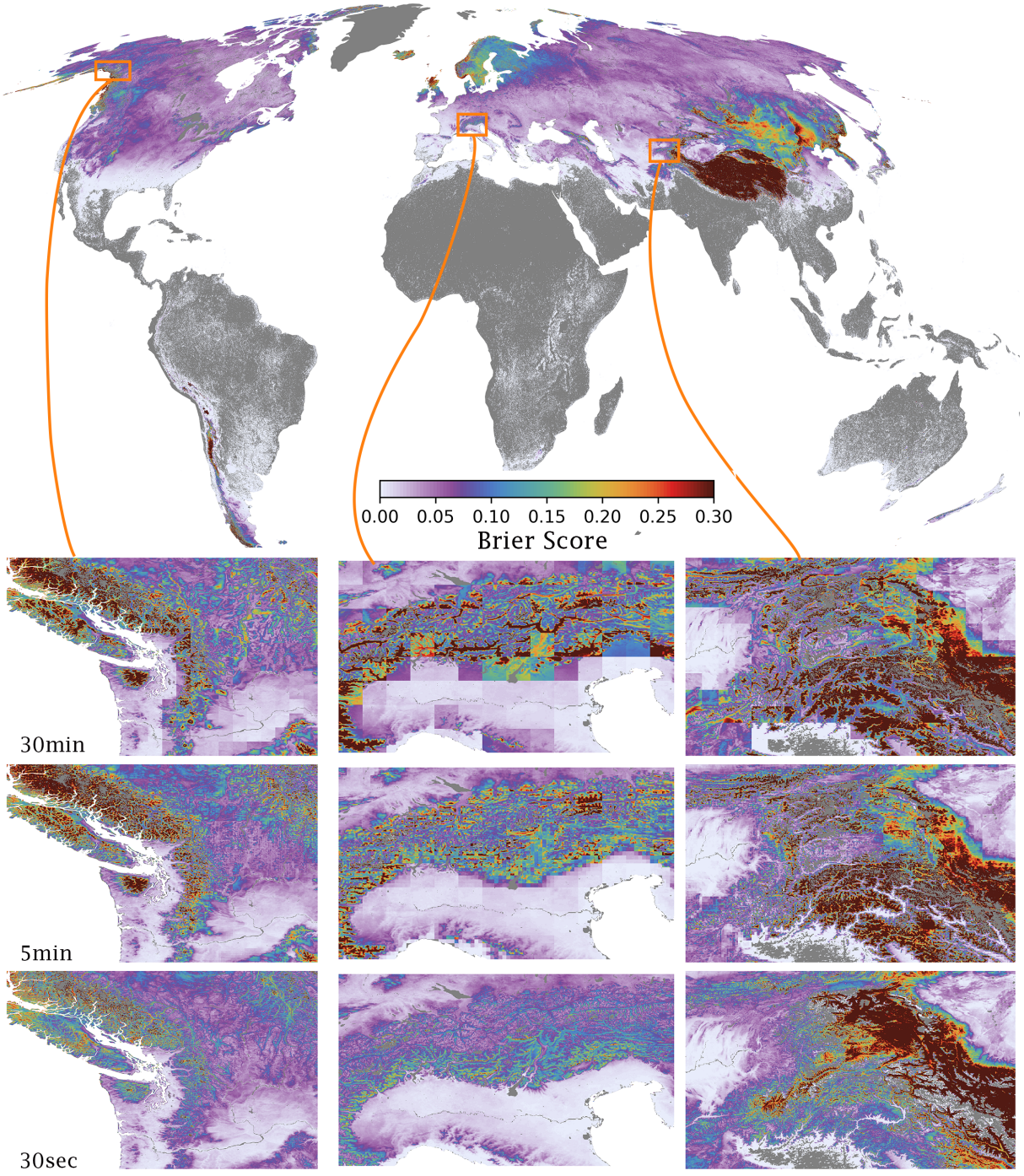

Figure 8. Brier score of simulated snow cover (-) simulated by PCR-GLOBWB at 30 arc-seconds resolution. Zoomed insets compare how the highlighted regions differ between the 30 arc-seconds, 5 arc-minutes and 30 arc-minutes resolution PCR-GLOBWB simulations.

3.5 River Discharge

Simulated river discharge more closely resembles the observations in the 30 arc-seconds resolution with close to 80% of the 7 086 stations showing skillful discharge simulations when compared to the 5 arc-minutes (70%) and the 30 arc-minutes (30%, Fig. 9a). The improvement in the 30 arc-second simulations is mainly brought about by increases in the correlation values and reductions in the bias and variance errors (Fig. 9b,c,d). The 30 arc-seconds simulation improves upon the 30 arc-minutes and 5 arc-minutes simulation in the majority of locations and this improvement is more pronounced for smaller catchments compared to larger catchments (Fig. 10a,b). The increase in performance related to resolution increase is applicable to all elevations and basins sizes, however the magnitude of performance increase is greatest for basins that are characterised by high elevations (Fig. 10c) and small catchment areas (Fig. 10d).

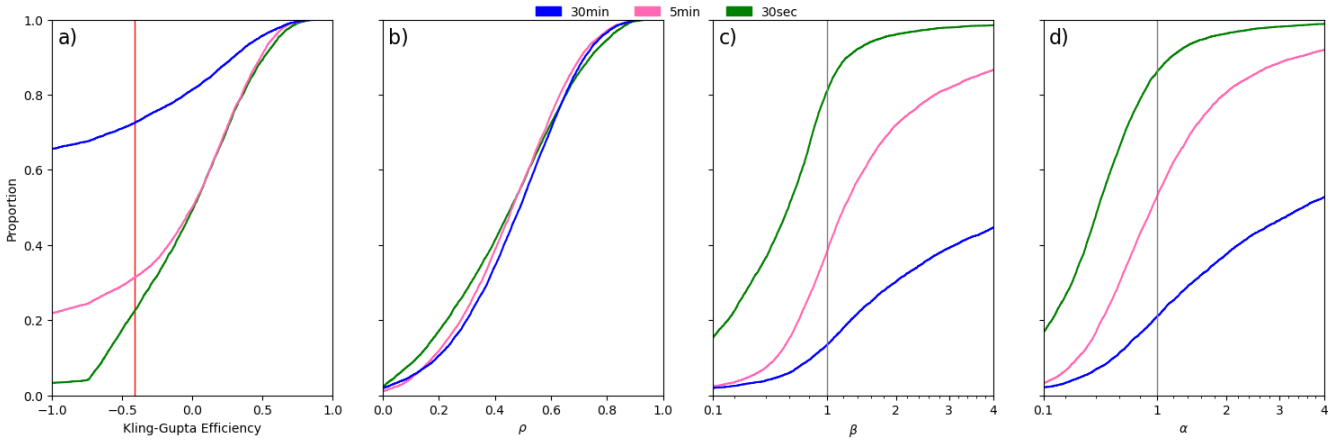

Figure 9. (a) KGE, (b) correlation component, (c) beta component, (d) alpha component calculated for daily river discharge from a 30 arc-minutes, 5 arc-minutes and 30 arc-seconds global PCR-GLOBWB simulation from 1985 - 2019. Values greater than -0.41 (red line) indicate the value at which stations improves upon the mean flow benchmark (Knoben et al., 2019).

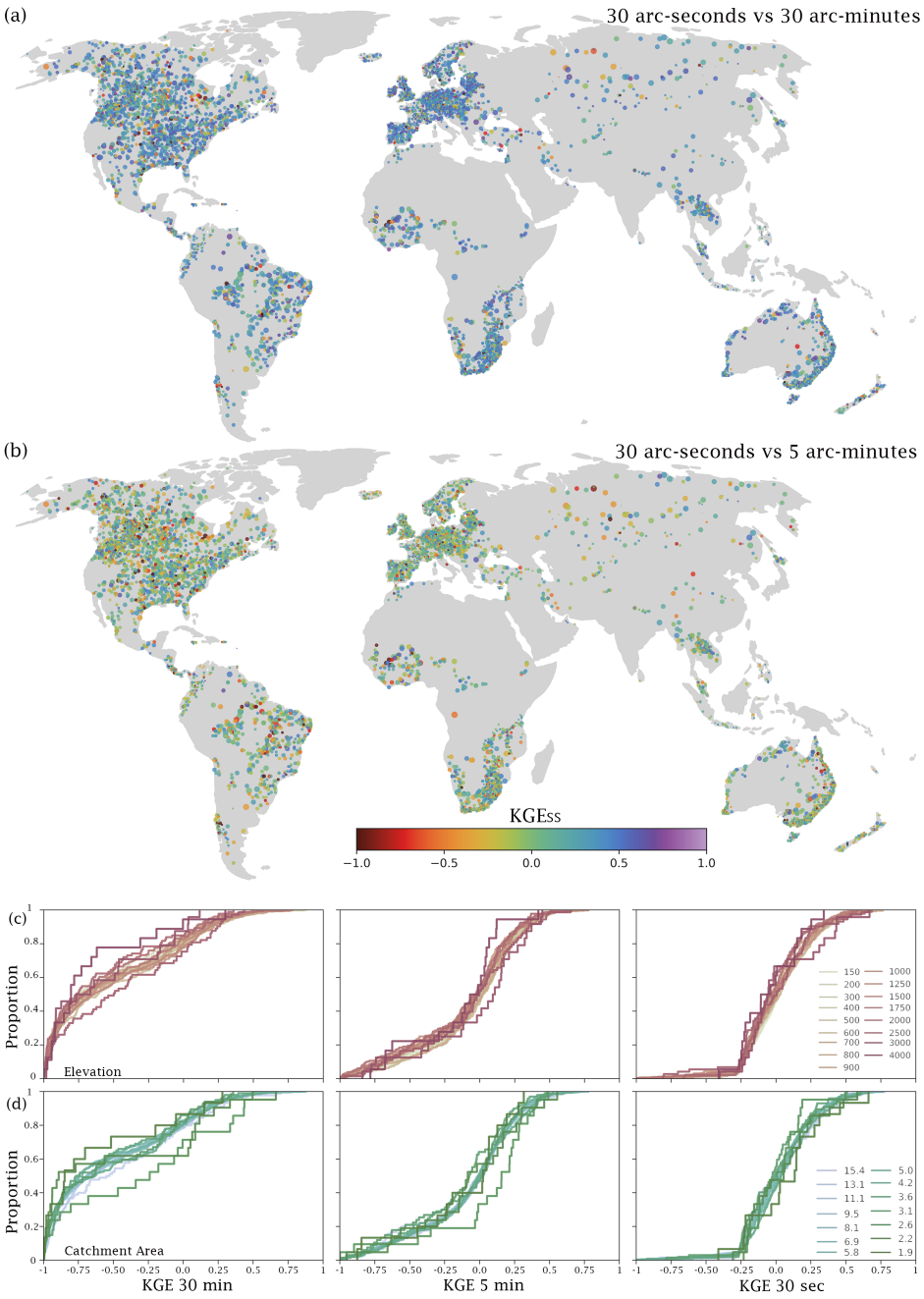

Figure 10. Spatial distribution of improvements in KGE skill score calculated for river discharge simulated using PCR-GLOBWB at the 30 arc-minutes, 5 arc-minutes and 30 arc-seconds resolution. (a) 30 arc-seconds vs 30 arc-minutes (b) 30 arc-seconds vs 5 arc-minutes. KGE cumulative distribution plots for catchments binned according to elevation (c) and catchment area (d).

4 Discussion

4.1 Computational and Storage demands

We show here that truly global hyper-resolution modelling is indeed possible with today's computational resources. However, long simulation times and excessive storage requirements for such simulations will hinder the reproducibility and accessibility for the wider hydrological community. Future attempts should consider how newer technologies may serve to reduce simulation times. For instance, relying on graphical processing units for computations has been shown to reduce simulation times (Freitas et al., 2022). Another opportunity lies in being able to enhance the parallelisation strategies used in global hydrological models. The simulations presented here employ a common method of parallelisation, termed 'embarrassingly parallel', where the modelling domain is subset into independent units (in this case basins) which are executed in parallel over multiple processes. Adopting alternative methods of parallelisation may further improve simulation times and a promising candidate called 'asynchronous many tasks' (AMT) may be especially useful to the hydrological community (de Jong et al., 2022). The use of an AMT based framework could not only improve simulation times but also reduce the ever increasing workload associated with post simulation data management. As an illustrative example, the simulation presented here produced $\sim$46 000 files spread across 645 folders, where as an AMT based simulation could result in as little as 72 files output files for the same simulation.

An important motivation for the development of global hyper-resolution models is that they are to be used in climate change studies (Wada et al., 2017); yet the storage and computational requirements for such an endeavour are still a limiting factor. To illustrate, here we evaluate four variables at a daily and one variable at the monthly frequency, which requires approximately 27 terabytes of storage. To obtain a more comprehensive view on the hydrological cycle, it would be preferable to have daily frequency data for more of the available model outputs. When considering all possible model outputs (114 variables) - this would yield a storage load of approximately one petabyte for a 35-year simulation. Extending this approach for a multi-model future simulation using the CMIP6 ensemble (approximately 130 models) storage requirements are estimated at 780 petabytes. Petascale storage requirements are inaccessible - or at the very least impractical - for most of the hydrological community. However, expected simulation times are less daunting, total simulation time would approach three months on the Dutch National Super Computer. Although this would provide valuable information, it also means that hydrology is now faced with the same issues current GCM's face; namely that, while such simulations are possible, data storage becomes a limitation (Schär et al., 2020). While solutions to the computational and storage limitations remain elusive; leveraging the recent trend towards cloud technologies for storing and disseminating earth science data is promising, yet the costs of these services remain high (Beven et al., 2015). To this end, the hyper-resolution hydrology community can benefit by emulating and drawing from the experiences of the remote sensing community who routinely depend on cloud computing and storage facilities to effectively disseminate large volumes of data to end users (Xu et al., 2022). Moreover, hyper-resolution physical based models may also benefit from adopting similar model deployment and dissemination strategies which are common amongst the machine learning community. For example, making intermediary states (ie., yearly states) available to the public could allow for more of the community to reproduce their required model outputs on the hardware available to them.

4.2 Increasing resolution and model accuracy

By increasing the model resolution we should expect to have a better representation of hydrological systems. However, we report mixed success when looking at the reproductions of different variables; below we discuss each of the evaluated variables.

Global Water Budget and Total Water Storage

Total water storage anomalies did not respond drastically to increasing model resolution, yet there were small reductions in the magnitude of error as modelling resolution is increased. This may be related to a more realistic distribution of water across the landscape at finer resolutions. Yet correlations were worse for the 30 arc-minutes, followed by the 30 arc-seconds and best for the 5 arc-minutes simulation. However, it is also important to note that benefits of higher resolutions may not be captured when using GRACE data to evaluate since the original resolution is at 30 arc-minutes and thus might by itself also capture a different signature than produced by the high-resolution simulations (Hoch et al., 2023). Partitioning between the major water reserves differed in response to increasing model resolution, with a markedly larger value of evaporation and lower discharge at the highest resolution. In comparison to other global models the 30 arc-seconds PCR-GLOBWB rates of evaporation are within previously reported ranges (60 000 - 85 000 $km^3 year^{-1}$) whereas runoff is slightly lower than previously reported (42 000 - 66 000 $km^3 year^{-1}$; Haddeland et al. (2011)). Interestingly, runoff is inline with a machine learning based estimate based off of station based river discharge data (Ghiggi et al., 2019). Expected evaporation to precipitation ratios are around 60% (Miralles et al., 2011), and for the 30 arc-minutes (57%) and 5 arc-minutes (56%) simulations this is the case. However, for the 30 arc-seconds resolution the evaporation to precipitation ratios is 69%, thus exceeding the expected value by approximately 9 percentage points.

Soil Moisture and Evaporation

Overall, predictions of soil moisture did not show much improvement as resolution increased, when considering the KGE scores. Differences do arise when the different components of the KGE scores are compared between the resolutions. The correlation and variability between observed and modelled soil moisture values show improvements, whilst there the magnitude of negative bias increase with increasiong resolution. It is important to note that the bias and variability is largely dependent on the accuracy of forcing data and differences between simulations across resolutions could be a result of forcing and not necessarily due to differences in the model. Nonetheless, the observation that the correlation and variability is better predicted and the bias is less well predicted as resolution increases could be explained by a scenario where the model overestimates evaporation, which in turn results in soils that are too dry. In congruence, the evaporation estimates at the 30 arc-seconds are significantly worse than those of the lower resolutions and tends towards an overestimation in the 30 arc-seconds when compared to 5 and 30 arc-minutes resolution. It is feasibly that this could be due to difference in how the meteorological forcing are scaled between the resolutions or the parameterization of the model that determines the rates of evaporation. As shown hereafter in section 4.3, it is likely due to the combination of parameterization and forcing effects.

Snow Cover

Increasing resolution resulted in better spatial representation of snow cover, with the highest resolution presenting the highest accuracy. Increased accuracy is brought about by the reduction in instances where snow is simulate in absence of snow in

observations a results directly related to increased resolution. Although differences are small when considering the global picture, improvements are most prominent in high elevation regions. These improvements are related to both increasing modelling resolution or the introduction of lateral snow transport which prevents the formation of erroneous snow towers. Although not the direct focus of this work, future studies should look at how snow can be better represented in global hydrological models by including process that are important in determining the water dynamics in ice and snow, especially when modelling at fine spatial resolutions. For instance, the current PCR-GLOBWB does not have unique glacier implementations. Yet, glaciers have been shown to locally and regionally important and including such processes do results in better predictions at larger scales (Hanus et al., 2024; Wiersma et al., 2022). Similarly, work needs to be done to improve snow dynamics and move beyond the simple snow melt model currently present in the model by including additionally processes which have also been shown to increase accuracy of predictions (Freudiger et al., 2017).

River Discharge

Predictions of river discharge improved markedly as modelling resolution increased. 30 arc-seconds displayed the most accurate predictions of river discharge. The improvement in KGE values at higher resolution is mostly the result of a better timing of the discharge peaks and troughs resulting in larger correlation coefficients (Fig. 10). Reductions in bias as resolution increases also contribute to improvement of river discharge; for the 30 arc-minutes and 5 arc-minutes resolution the model tends to overestimate river discharge, whilst the 30 arc-seconds results are underestimated. When considering that, for the 30 arc-seconds resolution, discharge values are underestimated and in conjunction with the observation that evaporation is overestimated, the question arises whether this increased evaporation leads to better estimates of river discharge by correcting for overestimation in the coarser resolution. Indeed from the soil moisture and evaporation validations, we can conclude that an overestimation of evaporation may result in a better estimation of river discharge bias. A positive result from these evaluations is that smaller catchments and catchments are higher elevations are now better represented by the model, a result directly related to increased model resolution. Increased resolutions are known to better represent smaller catchments (Hoch et al., 2023; O'Neill et al., 2021; Aerts et al., 2022).

In a broader sense a direct comparison with other global hydrological models are challenging given differences in validation approaches. Heinicke et al. (2024) reports that for 9 global hydrological models, the median KGE for daily river discharge ranges from -0.43 to 0.46. The KGE scores presented by Heinicke et al. (2024) were based of 644 stations and comparison to the 7 086 stations used in this study so direct comparison should be done with care, nonetheless it is encouraging to see that even with the differences in validation approaches, the 5 arc-minutes (median KGE=0.1) and 30 arc-seconds (median KGE=0.1) values are within range, even with a wider set of validation stations. Regarding other hyper-resolution hydrological models, the scores reported here are similar to that of wflow_sbm implemented over CONUS (median KGE=0.0) (Aerts et al., 2022).

4.3 Untangling Model Scaling and Forcing Downscaling Affects

The results above lend to the possibility the changes in forcing downscaling or parameteristion in 30 arc-seconds model may be responsible for the patterns we observe. Namely that overestimation in evaporation leads more accurate estimates of river discharge through correcting for an overestimation of river discharge in the 5 and 30 arc-minute simulation.

Table 4. Water balance components $km^3 year^{-1}$ for 1985 - 2019 for a 30 arc-minutes, 5 arc-minutes, and 30 arc-seconds simulated by PCR-GLOBWB over the European Continent. Values in parenthesis indicate the relative amounts in relation to precipitation as a percentage.

Resolution	Downscaling	Precipitation	Evaporation	Runoff	Storage
5 arc-minutes	Old	6 532	3 601 (55%)	2 804 (43%)	127
	New	6 937	4 002 (58%)	2 807 (40%)	127
30 arc-seconds	Old	6 377	4 468 (70%)	1 860 (30%)	49
	New	6 779	4 947 (73%)	1 787 (26%)	44

In order to differentiate between the effects of the downscaling methodology and those related to model parameterisation as resolution increases, the 5 arc-minutes and 30 arc-seconds simulations, using the old and new downscaling method, were compared over continental Europe and evaluated as above (Table 4 & Fig. 11). The fact that the new downscaling method provides comparable results when comparing it to the old downscaling method at the same resolution, suggests that the differences in the model results are most likely attributable to model parameterisation. One likely candidate is the way in which land cover is handled within the model at the 30 arc-second resolution. For the 30 arc-minute and 5 arc-minute models, PCR-GLOBWB allows for sub-grid variability in land cover type, where as in the 30 arc-seconds only the dominant land cover type is assumed per grid cell (Table 1). This was done to reduce computation time by avoiding having to loop over land cover classes for each time step. Since forest is often the predominant type, even in agricultural landscapes, and forests generally have higher evaporation than crops, evaporation is likely overestimated through an inflated representation of forests in the model. To corroborate this claim we conducted an post-hoc analysis which was aimed understanding which proportion of the model domain consists of forests when using dominant land cover types compared to a fractional coverage. This analysis revealed that by expressing land cover as a single dominant class per grid cell leads to a 13 percentage points (Appendix B1) inflation in the total area covered by forests ($\sim 50\%$) compared to when using the fractional cover ($\sim 37\%$).

To further evaluate the sensitivity of the water budget terms to changes in land cover parameterisation, for a small test region, we changed the land cover representation so that the entire region consisted of either forest, grasslands, or crops and compared the water budget terms to a 5 arc-minute simulation with unchanged land cover representation (see appendix B2). These simulations show that decreasing the relative abundance of forests within a domain will result in decreased rates of evaporation and increased rates of runoff. However although the results for this region are quite sensitive to land cover, it is unlikely that any combination of land covers will result in relative rates of evaporation and runoff similar to that of the coarser resolutions. Thus suggests that there are further opportunities, besides land cover representation, responsible for the difference

in water budgets between resolutions. For instance it may be that neither downscaling approach is capable of reproducing meteorology accurately at the 30 arc-seconds resolution.

Our results suggest that global hydrological models need to incorporate land cover heterogeneity even at the sub-kilometre scale in search of better predictive capacity. In congruence to this observation, land cover representation has previously been shown to be important in providing accurate predictions of hydrological states at the kilometre and even sub-kilometre resolution (Singh et al., 2015; Lazin et al., 2020). It is important to note that although needed for improving the accuracy of predictions, incorporating sub-kilometre land cover heterogeneity would further increase computation times. In addition, it is evident that the performance of high resolution horological models are hindered by the availability of accurate high resolution meteorological forcing, an observation that has previously been highlighted by more localised studies (Malle et al., 2024).

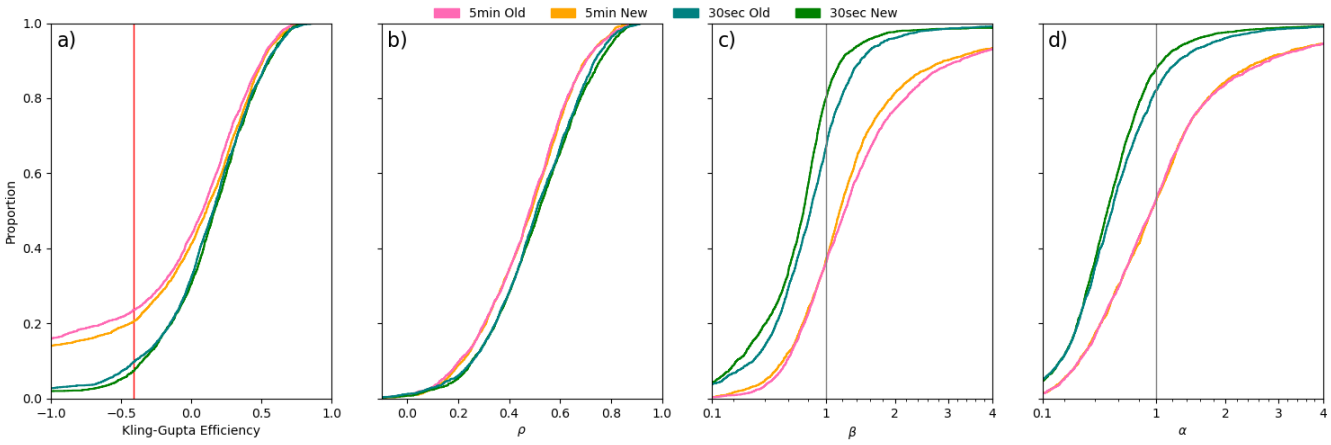

Figure 11. (a) KGE, (b) correlation component, (c) beta component, (d) alpha component calculated for daily river discharge from a 30 arc-minutes, 5 arc-minutes and 30 arc-seconds PCR-GLOBWB simulation from 1985 - 2019 for the European region. Values greater than -0.41 (red line) indicate the value at which stations improves upon the mean flow benchmark (Knoben et al., 2019).

4.4 Global Hyper-resolution hydrological modelling: status and recommendations

Our work advances global hyper-resolution modelling and its application beyond the continental scale (Hoch et al., 2023; O'Neill et al., 2021; Vergopolan et al., 2021; Chaney et al., 2021). We show that a hyper-resolution global hydrological model is feasible given the computational limits currently in place. However, it also highlights that storage is expected to become a significant challenge as global hydrological modelling advances. This will create additional challenges when output data are used for further analysis and when extracting point location time series in a computationally effective way.

The introduction of the climatology-centred downscaling method and ability to move frozen water stores laterally result in more continuous and ultimately a more realistic representation of the hydrological cycle. Hoch et al. (2023) proposed that enhancing the representation of additional physical processes at the kilometre scale could enhance the predictions. The findings presented in this study indicate that this is indeed true, although further efforts are required to further enhance the predictive

capabilities. The importance of high resolution forcing data is well understood and it has been shown that hydrological predictions become better when more high resolution data are used to force such models (Evin et al., 2024; Alfieri et al., 2022). Such data is only available for limited regions of the world and obtaining global and spatially coherent high resolution forcing data is a challenge that needs to be addressed if we are to have higher quality hydrological predictions at the global scale.

Furthermore, as also shown by (Hoch et al., 2023) there are a number of epistemic uncertainty issues related to global hydrological modelling that still need to be addressed. For instance we need a better representation of the land surface processes, snow and ice processes as these processes play a dominant role at fine spatial resolution and can no longer be neglected or captured in existing conceptual parametrization. At finer spatial resolution altitude effects start to play a key role in precipitation totals, snow and ice formation, melt and evaporation and thus require fine resolution meteorological information if possible at the global scale.

On the other hand, moving to a higher resolution allows for a better match with in situ observations and more recently released high resolution remote sensing products; the importance of scale commensurability between model outputs and that of in situ observations has been highlighted by Beven et al. (2022). For instance, the caravan dataset which has 6 830 stations for small river catchments (Kratzert et al., 2023), could be used to better underpin the accuracy of simulated river discharge values at higher resolution - as has been done for smaller scale studies (Aerts et al., 2023). Resolutions coarser than 30 arc-seconds would not allow for the inclusion of these river catchments, given the disconnect between modelling resolution and observed data station sizes. In addition, recent advances in remotely sensed high resolution soil moisture data will also be a valuable resource for evaluating hyper-resolution simulations once their time series are sufficiently long (Brocca et al., 2024).

5 Conclusions

The main goal of this study was to develop a unique global hyper-resolution hydrological model that covers a period of several decades (1985 – 2019), expanding the current hyper-resolution hydrological models beyond continental boundaries. We incorporated a novel downscaling approach and lateral movement of frozen water that ultimately yields more realistic representations of the hydrological cycle. Yet, as resolution increased the model tended to over-estimate rates of total evaporation which resulted in reductions in runoff. This suggests that additional processes that are relevant at the hyper resolution need further attention. Overall, the pursuit of hyper-resolution hydrological models are driven by the assumption that they will be able to provide stake-holders with more local estimates of water resources; one promising result reported in this study is that increased resolution is met with more accurate estimates of river discharge.

Appendix A: Spatial Distribution of Validation Stations

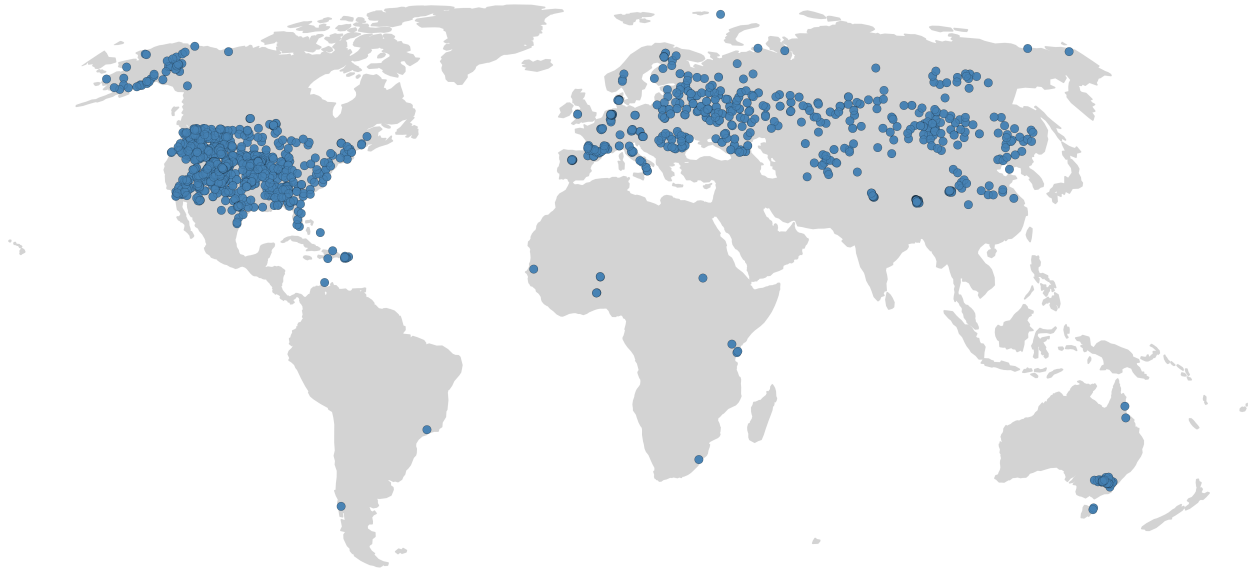

Figure A1. Map showing locations of stations used for validation of simulated soil moisture.

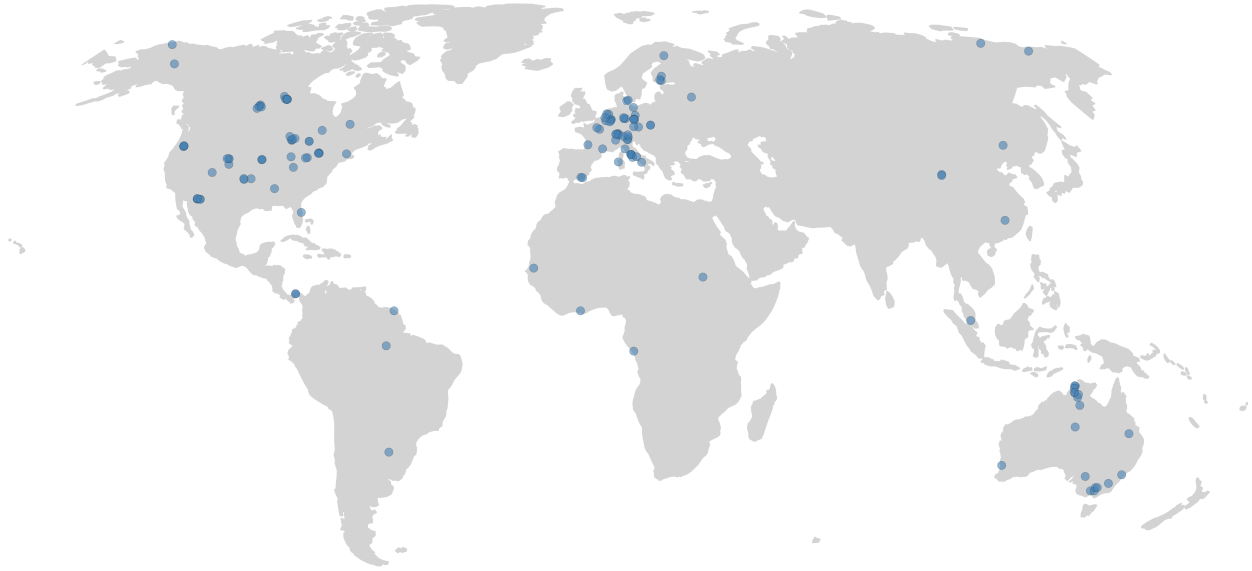

Figure A2. Map showing locations of stations used for validation of simulated total evaporation.

Appendix B: Land Cover Representation Analysis

B1 Dominant vs Fractional

The model evaluation shows that relative rates of land evaporation are higher in the 30 arc-seconds simulation in comparison with the 5 arc-minute and 30 arc-minutes simulations. We postulate that this is due to the way that the model represents land cover in the 30 arc-seconds simulation (Section 2.1 & 4.3). Given that the model we present here relies on the Global Land Cover Characteristics (GLCC) database version 2.0 (Loveland et al., 2000), which is at the 30 arc-seconds resolution, we can not directly test this assumption using the model input datasets. However, higher resolution land cover datasets are available and as such we leveraged these to test our hypothesis that using dominant land cover types leads to a higher proportion of the model domain being represented by forests than if fractional land cover types were used.

Here we rely on the Dynamic Land Cover map (100m resolution) from the Copernicus Global Land Service (Buchhorn et al., 2020) which provides both dominant land cover types and fractional land cover types at the 100m resolution. To test the difference between dominant and fractional, we compared to percentage of total land cover represented by forest when using the dominant land cover type compared to the fractional representation for the year 2019 (Buchhorn et al., 2020). When using the dominant land cover the total land area covered by forests amounts to 50% compared to when using fractional land cover this number is much lower at 37%, a difference of 13 percentage points.

B2 Water Budget Sensivity

In addition to the above analysis, we have also looked at how sensitive evaporation of the 30 arc-seconds simulation is to land cover changes, for a small test region (Southern Alps). We assumed that the entire domain is covered by either forest, grassland or croplands. These three simulations were compared to the original configuration 30 arc-seconds and 5 arc-minutes (as presented in the original manuscript).

Table B1. Water balance components $km^3 year^{-1}$ for 1985 - 2019 for a 5 arc-minutes, and 30 arc-seconds, with varying landcover parametrization, simulated by PCR-GLOBWB.

Simulation	Land cover	Precipitation	Evaporation	Runoff	Evaporation/Precipitation	Runoff/Precipitation
5 arc-minutes	not applicable	208	76	126	0.36	0.60
30 arc-seconds	not applicable	228	120	99	0.53	0.43
	Only Forest	228	130	89	0.57	0.39
	Only Grassland	228	124	96	0.54	0.42
	Only Crops	228	94	125	0.41	0.55

From this sensitivity test, we conclude that evaporation rates are indeed sensitive to changes in land cover representation, however no changes will allow us to reach the same evaporation rates as for the 5 arc-minutes or ratio between evaporation and precipitation. In addition the evaporation estimates we present are at the higher end of what can be expected (94 – 130

$km^3 year^{-1}$). It also shows that no land cover configuration at the 30 arc-seconds resolution will not reproduce the values of the 5 arc-minutes, since the maximum possible runoff values at the 30 arc-seconds (125 $km^3 year^{-1}$) are still lower than what is simulated by the 5 arc-minutes resolution (126 $km^3 year^{-1}$). This also means that even by changing all land cover to crops we can thus not achieve the same values as for the 5 arc-minutes model simulations.

Author contributions. **BvJ:** Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft preparation. **NW:** Conceptualization, Formal analysis, Project administration, Resources, Supervision, Writing – review & editing. **ES:** Conceptualization, Data curation, Software, Writing – review & editing. **JH:** Data curation, Software, Validation, Writing – review & editing. **BD:** Data curation, Software, Validation, Writing – review & editing. **JJ:** Formal analysis, Software, Writing – review & editing. **RvB:** Conceptualization, Data curation, Writing – review & editing. **MB:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Competing interests. The authors have also no other competing interests to declare.

Acknowledgements. This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative (using grant no. EINF-4875 & EINF-6855). MB acknowledges support from the ERC Advanced 969 Grant scheme (Grant no. 101019185 – GEOWAT). Some data and model parametrization stems from work done as part of the Climate-KIC project "Agriculture Resilient, Inclusive, and Sustainable Enterprise (ARISE)".

References

Aerts, J. P. M., Hut, R. W., van de Giesen, N. C., Drost, N., van Verseveld, W. J., Weerts, A. H., and Hazenberg, P.: Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model, Hydrology and Earth System Sciences, 26, 4407–4430, https://doi.org/10.5194/hess-26-4407-2022, 2022.

Aerts, J. P. M., Hoch, J. M., Coxon, G., van de Giesen, N. C., and Hut, R. W.: On the importance of observation uncertainty when evaluating and comparing models: a hydrological example, EGUsphere, pp. 1–26, https://doi.org/10.5194/egusphere-2023-1156, 2023.

Alcamo, J.: Water quality and its interlinkages with the Sustainable Development Goals, Current Opinion in Environmental Sustainability, 36, 126–140, https://doi.org/10.1016/j.cosust.2018.11.005, 2019.

Alfieri, L., Avanzi, F., Delogu, F., Gabellani, S., Bruno, G., Campo, L., Libertino, A., Massari, C., Tarpanelli, A., Rains, D., Miralles, D. G., Quast, R., Vreugdenhil, M., Wu, H., and Brocca, L.: High-resolution satellite products improve hydrological modeling in northern Italy, Hydrology and Earth System Sciences, 26, 3921–3939, https://doi.org/10.5194/hess-26-3921-2022, 2022.

Allen, R. G. and Food and Agriculture Organization of the United Nations, eds.: Crop evapotranspiration: guidelines for computing crop water requirements, no. 56 in FAO irrigation and drainage paper, Food and Agriculture Organization of the United Nations, Rome, ISBN 9789251042199, 1998.

Balland, V. and Arp, P. A.: Modeling soil thermal conductivities over a wide range of conditions, Journal of Environmental Engineering and Science, 4, 549–558, https://doi.org/10.1139/s05-007, 2005.

Benedict, I., van Heerwaarden, C. C., Weerts, A. H., and Hazeleger, W.: The benefits of spatial resolution increase in global simulations of the hydrological cycle evaluated for the Rhine and Mississippi basins, Hydrology and Earth System Sciences, 23, 1779–1800, https://doi.org/10.5194/hess-23-1779-2019, 2019.

Beven, K., Cloke, H., Pappenberger, F., Lamb, R., and Hunter, N.: Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface, Science China Earth Sciences, 58, 25–35, https://doi.org/10.1007/s11430-014-5003-4, 2015.

Beven, K., Lane, S., Page, T., Kretzschmar, A., Hankin, B., Smith, P., and Chappell, N.: On (in)validating environmental models. 2. Implementation of a Turing-like test to modelling hydrological processes, Hydrological Processes, 36, e14 703, https://doi.org/10.1002/hyp.14703, 2022.

Beven, K. J. and Cloke, H. L.: Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al., Water Resources Research, 48, 2011WR010 982, https://doi.org/10.1029/2011WR010982, 2012.

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

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., De Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., Van De Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?: "Everywhere and locally relevant", Hydrological Processes, 29, 310–320, https://doi.org/10.1002/hyp.10391, 2015.

Brocca, L., Gaona, J., Bavera, D., Fioravanti, G., Puca, S., Ciabatta, L., Filippucci, P., Mosaffa, H., Esposito, G., Roberto, N., Dari, J., Vreugdenhil, M., and Wagner, W.: Exploring the Actual Spatial Resolution of 1 Km Satellite Soil Moisture Products, https://doi.org/10.2139/ssrn.4737858, 2024.

Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., and Karger, D. N.: CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution, https://doi.org/10.16904/ENVIDAT.332, 2022.

Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe, https://doi.org/10.5281/zenodo.3939050, 2020.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., and Wada, Y.: Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management, Geoscientific Model Development, 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020, 2020.

Chaney, N. W., Torres-Rojas, L., Vergopolan, N., and Fisher, C. K.: HydroBlocks v0.2: enabling a field-scale two-way coupling between the land surface and river networks in Earth system models, Geoscientific Model Development, 14, 6813–6832, https://doi.org/10.5194/gmd-14-6813-2021, 2021.

de Jong, K., Panja, D., Karssenberg, D., and van Kreveld, M.: Scalability and composability of flow accumulation algorithms based on asynchronous many-tasks, Computers & Geosciences, 162, 105 083, https://doi.org/10.1016/j.cageo.2022.105083, 2022.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M., Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrology and Earth System Sciences, 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., and Wada, Y.: Modelling Freshwater Resources at the Global Scale: Challenges and Prospects, Surveys in Geophysics, 37, 195–221, https://doi.org/10.1007/s10712-015-9343-1, 2016.

Evin, G., Le Lay, M., Fouchier, C., Penot, D., Colleoni, F., Mas, A., Garambois, P.-A., and Laurantin, O.: Evaluation of hydrological models on small mountainous catchments: impact of the meteorological forcings, Hydrology and Earth System Sciences, 28, 261–281, https://doi.org/10.5194/hess-28-261-2024, 2024.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

Freitas, H. R. A., Mendes, C. L., and Ilic, A.: Performance optimization of the MGB hydrological model for multi-core and GPU architectures, Environmental Modelling & Software, 148, 105 271, https://doi.org/10.1016/j.envsoft.2021.105271, 2022.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., and Weiler, M.: Snow redistribution for the hydrological modeling of alpine catchments, WIREs Water, 4, e1232, https://doi.org/10.1002/wat2.1232, 2017.

Frey, S. and Holzmann, H.: A conceptual, distributed snow redistribution model, Hydrology and Earth System Sciences, 19, 4517–4530, https://doi.org/10.5194/hess-19-4517-2015, 2015.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, Earth System Science Data, 11, 1655–1674, https://doi.org/10.5194/essd-11-1655-2019, 2019.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, Journal of Hydrometeorology, 12, 869–884, https://doi.org/10.1175/2011JHM1324.1, 2011.

Hagemann, S.: An improved land surface parameter dataset for global and regional climate models, Tech. Rep. 336, Max-Planck-Institut für Meteorologie, 10.17617/2.2344576, 2002.

Hagemann, S. and Gates, L. D.: Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations, Climate Dynamics, 21, 349–359, https://doi.org/10.1007/s00382-003-0349-x, 2003.

Hagemann, S., Botzet, M., Dümenil, L., and Machenhauer, B.: Derivation of global GCM boundary conditions from 1 km land use satellite data, Tech. Rep. 289, Max-Planck-Institut für Meteorologie, https://pure.mpg.de/pubman/faces/ViewItemFullPage.jsp?itemId=item_1562156_5, 1999.

Hallouin, T.: hydroeval: an evaluator for streamflow time series in Python, https://doi.org/10.5281/zenodo.4709652, 2021.

Hanus, S., Schuster, L., Burek, P., Maussion, F., Wada, Y., and Viviroli, D.: Coupling a large-scale glacier and hydrological model (OGGM v1.5.3 and CWatM V1.08) – Towards an improved representation of mountain water resources in global assessments, EGUsphere, pp. 1–31, https://doi.org/10.5194/egusphere-2023-2562, 2024.

Heinicke, S., Volkholz, J., Schewe, J., Gosling, S. N., Schmied, H. M., Zimmermann, S., Mengel, M., Sauer, I. J., Burek, P., Chang, J., Kou-Giesbrecht, S., Grillakis, M., Guillaumot, L., Hanasaki, N., Koutroulis, A., Otta, K., Qi, W., Satoh, Y., Stacke, T., Yokohata, T., and Frieler, K.: Global hydrological models continue to overestimate river discharge, Environmental Research Letters, 19, 074 005, https://doi.org/10.1088/1748-9326/ad52b0, 2024.

Hengl, T., Jesus, J. M. d., Heuvelink, G. B. M., Gonzalez, M. R., 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., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, e0169 748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., van Beek, R. L. P. H., and Bierkens, M. F. P.: Hyper-resolution PCR-GLOBWB: opportunities and challenges from refining model spatial resolution to 1 km over the European continent, Hydrology and Earth System Sciences, 27, 1383–1401, https://doi.org/10.5194/hess-27-1383-2023, 2023.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, Scientific Data, 4, 170 122, https://doi.org/10.1038/sdata.2017.122, 2017.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Kornfeld, R. P., Arnold, B. W., Gross, M. A., Dahya, N. T., Klipstein, W. M., Gath, P. F., and Bettadpur, S.: GRACE-FO: The Gravity Recovery and Climate Experiment Follow-On Mission, Journal of Spacecraft and Rockets, 56, 931–951, https://doi.org/10.2514/1.A34326, 2019.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Scientific Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Müller Schmied, H., Hersbach, H., Buontempo, C., and Cagnazzo, C.: WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0), https://doi.org/10.48364/ISIMIP.342217, type: dataset, 2021.

Lazin, R., Shen, X., Koukoula, M., and Anagnostou, E.: Evaluation of the Hyper-Resolution Model-Derived Water Cycle Components Over the Upper Blue Nile Basin, Journal of Hydrology, 590, 125 231, https://doi.org/10.1016/j.jhydrol.2020.125231, 2020.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, https://doi.org/10.1016/j.jhydrol.2004.03.028, 2004.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrological Processes, 27, 2171–2186, https://doi.org/10.1002/hyp.9740, 2013.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos, Transactions American Geophysical Union, 89, 93–94, https://doi.org/10.1029/2008EO100001, 2008.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494–502, https://doi.org/10.1890/100125, 2011.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303–1330, https://doi.org/10.1080/014311600210191, 2000.

Malle, J. T., Mazzotti, G., Karger, D. N., and Jonas, T.: Regionally optimized high-resolution input datasets enhance the representation of snow cover in CLM5, Earth System Dynamics, 15, 1073–1115, https://doi.org/10.5194/esd-15-1073-2024, 2024.

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

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrology and Earth System Sciences, 18, 3511–3538, https://doi.org/10.5194/hess-18-3511-2014, 2014.

Nagler, T., Schwaizer, G., Keuris, L., Hetzenecker, M., and Metsämäki, S.: ESA Snow Climate Change Initiative (Snow_cci): Daily global Snow Cover Fraction - snow on ground (SCFG) from MODIS (2000-2019), version 1.0, https://doi.org/10.5285/3B3FD2DAF3D34C1BB4A09EFEAF3B8EA9, 2021.

Olson, J.: Global ecosystem framework-translation strategy, Tech rep., USGS EROS Data Center Internal Report, Sioux Falls, SD, 1994a.

Olson, J.: Global ecosystem framework-definitions, Tech. rep., USGS EROS Data Center Internal Report, USGS EROS Data Center Internal Report, 1994b.

O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M.: Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States, Geoscientific Model Development, 14, 7223–7254, https://doi.org/10.5194/gmd-14-7223-2021, 2021.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. d., Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. d., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B.,

Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, , Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V. R. d., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Girolamo, S. D., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T. C., Sprenger, M., Ubbiali, S., and Wernli, H.: Kilometer-Scale Climate Models: Prospects and Challenges, Bulletin of the American Meteorological Society, 101, E567–E587, https://doi.org/10.1175/BAMS-D-18-0167.1, 2020.

Singh, R. S., Reager, J. T., Miller, N. L., and Famiglietti, J. S.: Toward hyper-resolution land-surface modeling: The effects of fine-scale topography and soil texture on CLM 4.0 simulations over the S outhwestern U.S., Water Resources Research, 51, 2648–2667, https://doi.org/10.1002/2014WR015686, 2015.

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, Hydrology and Earth System Sciences, 15, 2913–2935, https://doi.org/10.5194/hess-15-2913-2011, 2011.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geoscientific Model Development, 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Teluguntla, P., Thenkabail, P., Xiong, J., Gumma, M., Giri, C., Milesi, C., Ozdogan, M., Congalton, R., Tilton, J., Sankey, T., Massey, R., Phalke, A., and Yadav, K.: NASA Making Earth System Data Records for Use in Research Envi-

ronments (MEaSUREs) Global Food Security Support Analysis Data (GFSAD) Crop Mask 2010 Global 1 km V001, https://doi.org/10.5067/MEASURES/GFSAD/GFSAD1KCM.001, 2016.

Todini, E.: The ARNO rainfall—runoff model, Journal of Hydrology, 175, 339–382, https://doi.org/10.1016/S0022-1694(96)80016-3, 1996.

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., Coughlan de Perez, E., and Stephens, E. M.: Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin, Hydrology and Earth System Sciences, 23, 3057–3080, https://doi.org/10.5194/hess-23-3057-2019, 2019.

van Beek, L. P. H.: Forcing PCR-GLOWB with CRU data, Tech. rep., Department of Physical Geography, Utrecht University,, https://vanbeek.geo.uu.nl/suppinfo/vanbeek2008.pdf, 2008.

van Beek, L. P. H. and Bierkens, M. F. P.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, and Verification, Tech. rep., Department of Physical Geography, Utrecht University,, http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2008.

Van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resources Research, 47, 2010WR009 791, https://doi.org/10.1029/2010WR009791, 2011.

Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., Torres-Rojas, L., Sadri, S., and Wood, E. F.: SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US, Scientific Data, 8, 264, https://doi.org/10.1038/s41597-021-01050-2, 2021.

Vremec, M. and Collenteur, R.: PyEt - a Python package to estimate potential and reference evapotranspiration, Tech. Rep. EGU21-15008, Copernicus Meetings, https://doi.org/10.5194/egusphere-egu21-15008, 2021.

Vörösmarty, C. J., Hoekstra, A. Y., Bunn, S. E., Conway, D., and Gupta, J.: Fresh water goes global, Science, 349, 478–479, https://doi.org/10.1126/science.aac6009, 2015.

Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T. J., van Dijk, A. I. J. M., van Emmerik, T., Van Huijgevoort, M. H. J., Van Lanen, H. A. J., Vörösmarty, C. J., Wanders, N., and Wheater, H.: Human–water interface in hydrological modelling: current status and future directions, Hydrology and Earth System Sciences, 21, 4169–4193, https://doi.org/10.5194/hess-21-4169-2017, 2017.

Wiersma, P., Aerts, J., Zekollari, H., Hrachowitz, M., Drost, N., Huss, M., Sutanudjaja, E. H., and Hut, R.: Coupling a global glacier model to a global hydrological model prevents underestimation of glacier runoff, Hydrology and Earth System Sciences, 26, 5971–5986, https://doi.org/10.5194/hess-26-5971-2022, 2022.

Wilby, R. L., Hay, L. E., Gutowski, W. J., Arritt, R. W., Takle, E. S., Pan, Z., Leavesley, G. H., and Clark, M. P.: Hydrological responses to dynamically and statistically downscaled climate model output, Geophysical Research Letters, 27, 1199–1202, https://doi.org/10.1029/1999GL006078, 2000.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., De Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., Van De Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resources Research, 47, 2010WR010 090, https://doi.org/10.1029/2010WR010090, 2011.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., Liu, J., Wang, T., Yan, Z., Zhu, J., Jiang, T., and Guo, H.: Cloud-based storage and computing for remote sensing big data: a technical review, International Journal of Digital Earth, 15, 1417–1445, https://doi.org/10.1080/17538947.2022.2115567, 2022.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resources Research, 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.

Yang, C., Tijerina-Kreuzer, D., Tran, H., Condon, L., and Maxwell, R.: A high-resolution, 3D groundwater-surface water simulation of the contiguous US: Advances in the integrated ParFlow CONUS 2.0 modeling platform, https://eartharxiv.org/repository/view/5670/, 2023.