



HyIR 0.1 (HYdrological and Industrial Resilience): A Global Hydrological Aggregation Framework for Translating CMIP6 Climate Projections into Freshwater Availability Scenarios

Mariam Taki^{1,★}, Nicolas Flipo^{1,★}, and Damien Goetz^{1,★}

★These authors contributed equally to this work.¹Department of Geosciences and Geoengineering, Mines Paris – PSL University, 77300 Fontainebleau, France

Correspondence: Mariam Taki (mariam.taki@minesparis.psl.eu)

Abstract. Climate change is altering the global water cycle with direct consequences for water availability and the resilience of human activities. To support prospective assessment of freshwater availability under contrasting socioeconomic trajectories, we introduce HyIR (HYdrological and Industrial Resilience), a global-scale hydrological aggregation framework designed to translate CMIP6 climate projections into scenario-consistent estimates of freshwater storage and fluxes across key freshwater availability compartments.

This paper presents the hydrological aggregator of HyIR 0.1, forced by CMIP6 precipitation, evapotranspiration, and runoff fields covering the historical period (1850–2014) and future Shared Socioeconomic Pathways (2015–2100). The framework propagates these forcings through three coupled storage and routing components: surface water body storage, groundwater storage, and river network routing, grounded in global geophysical datasets. It provides explicit estimates of water stock evolution and exchanges between rivers, lakes and reservoirs, and groundwater storage compartments. A validation against observed daily river discharge from the Global Runoff Data Centre and the GloFAS reanalysis is presented for a proof-of-concept application to the TaiESM1 CMIP6 model. HyIR is intended as a decision-support tool for long-term freshwater availability assessment, with particular relevance to water-dependent industries. In contrast to many climate impact studies that report river discharge as the primary indicator, HyIR provides source-specific indicators with explicit storage accounting across rivers, lakes and reservoirs, and groundwater storage compartments.

1 Introduction

Anthropogenic climate change is reshaping the global water cycle with measurable and accelerating consequences. Rising surface temperatures driven by greenhouse gas emissions are intensifying evapotranspiration, altering precipitation patterns, and accelerating the melting of glaciers and snowpack, all of which affect the timing and magnitude of freshwater flows and storage (Huntington, 2006; Schewe et al., 2014; Haddeland et al., 2014). The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6) concludes with high confidence that water insecurity will intensify across large portions of the globe under all Shared Socioeconomic Pathways (SSPs), with projected increases in the frequency and severity of droughts, declining groundwater recharge, and growing mismatches between water supply and demand (Masson-Delmotte



et al., 2021; Seneviratne et al., 2021). While the implications of these projections for mean air temperature and atmospheric
CO₂ concentrations are now well constrained by the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al.,
2016), translating these climate signals into physically interpretable changes in freshwater availability in rivers, lakes, and
groundwater storage compartments remains a fundamental challenge. Such translation is an operational requirement for the
design of adaptation and resilience strategies, particularly for sectors whose viability depends directly on reliable access to
freshwater, including energy production, agriculture, and the extractive industries (Vörösmarty et al., 2010; Rockström et al.,
2009; United Nations World Water Assessment Programme (WWAP), 2019).

Global hydrological models (GHMs) have been developed and continuously improved since the mid-1990s to provide quan-
titative estimates of water flows and storages (Bierkens, 2015). Pioneering and widely used frameworks include WaterGAP
(Döll et al., 2003; Müller Schmied et al., 2021), PCR-GLOBWB (Sutanudjaja et al., 2018), VIC (Liang et al., 1994), and mHM
(Samaniego et al., 2010), among others. These models have been applied to a broad range of purposes: quantifying global water
stress and scarcity (Alcamo et al., 2007; Döll et al., 2018; Veldkamp et al., 2018), assessing flood hazard (Hirabayashi et al.,
2013), evaluating the hydrological impacts of land use and irrigation (Sacks et al., 2010), and contributing to international
multi-model comparison exercises such as the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Frieler et al.,
2017). The primary focus of most GHMs has been on total renewable water resources and the quantification of water avail-
ability expressed as river discharge, with increasing attention to human water use and man-made reservoirs as drivers of water
stress (Müller Schmied et al., 2021; Wada and Bierkens, 2014). These models typically operate with their own meteorological
forcing pipelines, based on gridded observational or reanalysis products (Müller Schmied et al., 2021), and have historically
been driven by climate projections from earlier CMIP generations. The representation of groundwater in most GHMs remains
conceptually simplified, generally treating subsurface storage as a lumped linear reservoir rather than a spatially differentiated
aquifer system (Taylor et al., 2013; Bierkens, 2015), which limits the ability to resolve heterogeneous aquifer responses or
to identify water available specifically from groundwater versus surface water compartments. Furthermore, most global-scale
climate impact assessments ultimately report river discharge as the primary indicator of water availability. Most GHMs there-
fore report water availability as aggregate river discharge at the basin or grid-cell scale, providing only limited information on
groundwater stocks. The distinction between extractable water by source type, whether from lakes and wetlands, river reaches,
or groundwater storage, at the spatial granularity required for site-level industrial applications, remains largely unaddressed in
existing global frameworks.

These considerations motivate the development of a framework operating directly on the latest generation of climate pro-
jections from CMIP6. In HyIR, precipitation, evapotranspiration, and runoff produced by each GCM realization are taken
as boundary conditions and their partitioning is preserved. HyIR then translates these GCM outputs into scenario-consistent,
physically interpretable estimates of freshwater availability across compartments: routed river discharge, surface water body
storage in lakes, reservoirs and wetlands, and a groundwater storage compartment. Such a framework should cover the full
CMIP6 temporal envelope from the historical period through 2100 under multiple SSPs, and provide spatially explicit, source-
specific water availability indicators. It provides a consistent aggregation of GCM-derived fields within a unified topological
and storage representation, enabling systematic analysis of freshwater availability across the CMIP6 ensemble (Gosling et al.,



2017; Bierkens, 2015; Haddeland et al., 2014). The intended application extends beyond scientific water cycle characteriza-
tion to operational water-risk assessment for industrial sectors whose infrastructure is exposed to climate-driven changes in
freshwater availability.

This paper presents the hydrological aggregator of HyIR (HYdrological and Industrial Resilience), a global open-source
platform designed to assess freshwater availability and water-related risks for industrial operations under climate change. The
full HyIR framework spans from climate forcing to site-level industrial water risk: starting from CMIP6 climate projections, it
produces spatially explicit estimates of water availability by source, which are subsequently used to assess exposure and access
to water for georeferenced industrial facilities, support the evaluation of water-sourcing strategies, and inform adaptation
planning at the global scale. The present paper focuses exclusively on the hydrological aggregator, which constitutes the
physical backbone of this broader framework.

HyIR 0.1 is a global-scale hydrological aggregation framework that leverages CMIP6 precipitation, evapotranspiration, and
runoff fields to reconstruct scenario-consistent water stocks and fluxes relevant for freshwater availability assessments, over
the historical period (1850–2014) and future SSP scenarios (2015–2100). These climate fields are propagated through three
coupled components: a surface water body storage component, a groundwater storage component, and a river network routing
scheme, operating on the native spatial grid of each GCM. HyIR should not be interpreted as a global hydrological model:
it is not prognostic and does not independently simulate land surface or hydrogeological processes beyond what is needed
to translate climate projections into source-differentiated water availability indicators. Its distinguishing characteristic is the
explicit representation of water storage by compartment, including a spatially differentiated groundwater storage scheme based
on global depth-to-bedrock data, allowing availability to be estimated separately in rivers, lakes and reservoirs, and groundwater
storage compartments. The primary application target is the mining industry, which concentrates important freshwater demands
in certain arid and water stressed regions (Northey et al., 2017; Mudd, 2010; International Council on Mining and Metals
(ICMM), 2012). As a proof of concept, this paper presents the model description and a validation of the hydrological aggregator
against observed river discharge, followed by an application to historical and future projections using the TaiESM1 (Taiwanese
Earth System model) CMIP6 general circulation model.

The paper is structured as follows. Section 2 presents the model description, detailing the components and governing equa-
tions of the HyIR hydrological aggregator. Section 3 describes the evaluation framework, covering the observational datasets,
matching methodology, and performance metrics. Section 4 presents the results of an application of the framework forced
by the TaiESM1 GCM, including calibration and historical validation results. Section 5 discusses framework applications,
limitations, and future developments. Section 6 concludes.



2 Model description

2.1 Overview

90 The HyIR hydrological aggregator is a global-scale aggregation framework forced by monthly outputs from CMIP6 GCMs. It leverages GCM-provided precipitation, evapotranspiration, and runoff as boundary conditions. These forcing fields are redistributed through three coupled storage and routing components (Fig. 1):

1. **Surface water body storage:** computes the water balance of individual lakes and reservoirs using polygons from the Global Lakes and Wetlands Database (GLWD);
- 95 2. **Groundwater storage:** represents groundwater storage units delineated from global depth-to-bedrock data;
3. **River network routing:** propagates lateral inflow through the HydroRIVERS network using a semi-implicit Muskingum scheme at a daily sub-timestep.

The spatial backbone is the native grid of the selected GCM. Each continental land grid cell is referred to as a `HydroUnit` and constitutes the fundamental unit of the water balance. No soil moisture layer or snow module is included; all land surface partitioning is inherited directly from the driving GCM.

100 Within each monthly time step, compartments are solved sequentially. Lake storage is updated first using the previous month's volumes and the current month's CMIP6 forcing, producing lake-groundwater and lake-river exchange fluxes for time step t . These then feed the groundwater balance at the same step t , yielding baseflow contributions to the river network. The resulting total lateral inflow to each river reach is assembled and propagated at a daily sub-timestep using a Muskingum routing scheme over the corresponding number of days in the month. All state variables from month t therefore serve as initial conditions for month $t + 1$.

2.2 Spatial discretization and regional decomposition

2.2.1 HydroUnit grid

110 The computational mesh of HyIR follows the native grid of the selected CMIP6 GCM. The spatial resolution of CMIP6 models ranges from approximately $0.5^\circ \times 0.5^\circ$ to $2.8125^\circ \times 2.8125^\circ$ (ESGF, 2026). Each continental land cell, defined as any grid cell containing a non-zero land fraction, is retained as a `HydroUnit`. Purely oceanic cells and cells corresponding to Antarctica are excluded; the latter falls outside the scope of HyIR, as freshwater availability and industrial water demand are not relevant in that region. `HydroUnit` land areas A_{HU} are computed in an equal-area projection (EPSG:6933).

2.2.2 Continental delineation and parallelization

115 River routing in HyIR requires knowledge of the upstream-downstream connectivity of every river reach, which introduces a global topological dependency. To make global-scale runs computationally tractable while preserving physical correctness, the

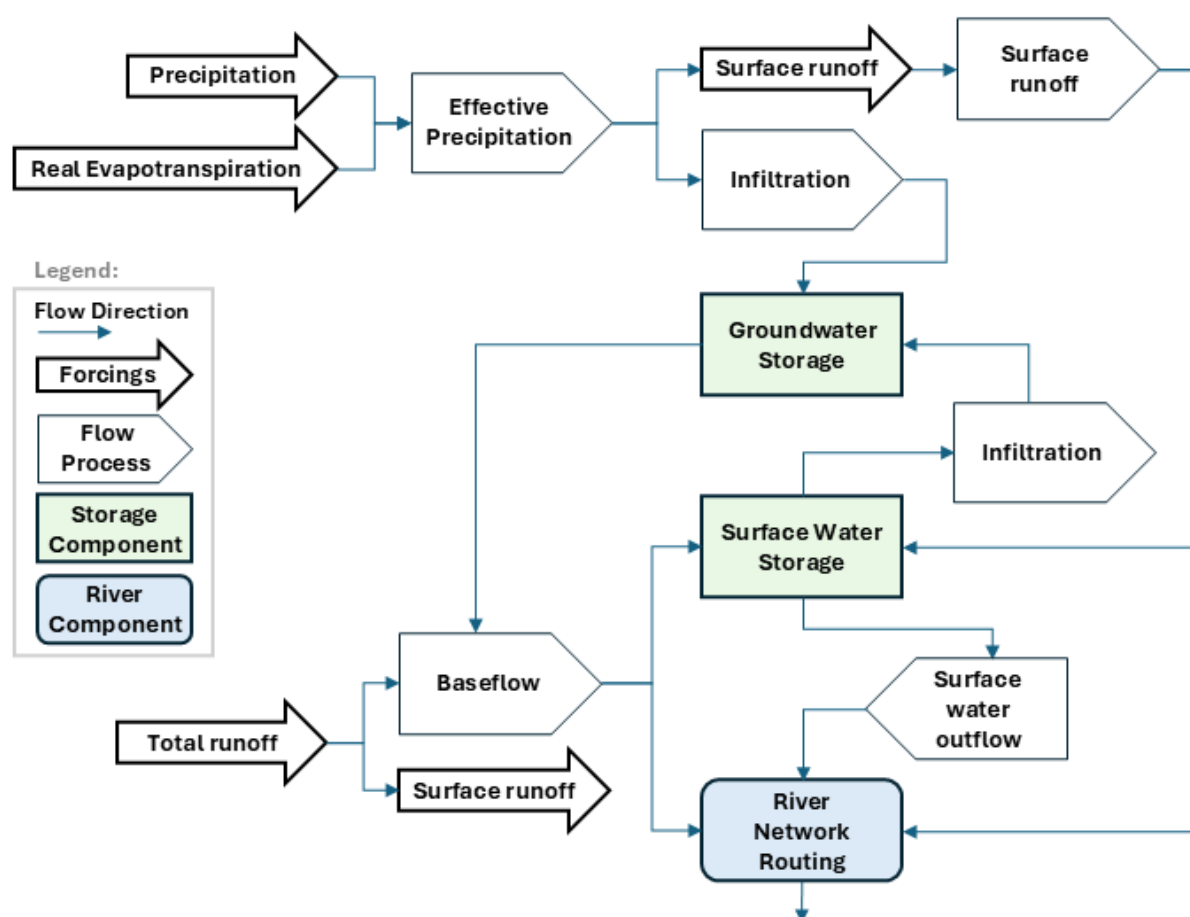


Figure 1. Schematic overview of the HyIR hydrological aggregator. Climate forcings enter the system at the HydroUnit level and are partitioned among lakes and wetlands, the groundwater storage compartment, and the river network. Arrows indicate the direction of water flux.

global river network is partitioned into nine independent continental domains following the Pfafstetter Level 1 boundaries of HydroBASINS (Lehner and Grill, 2013): Africa, Asia, Arctic, Australia, Europe, Greenland, North America, South America, and Siberia (Fig. 2). These regions are hydrologically independent in terms of surface drainage: no major river system crosses their boundaries, so routing can be performed in parallel across domains without loss of mass or connectivity.

The full hydrology pipeline, including water balance, lake storage, groundwater, and river routing, is executed independently for each of these nine domains, enabling parallel computation across regions on multi-core environments.

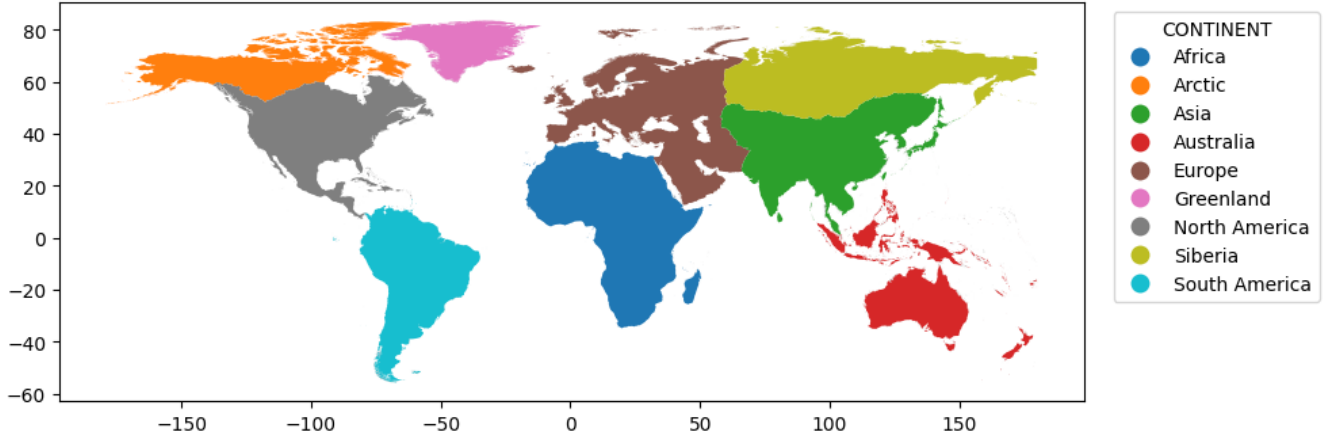


Figure 2. The nine continental domains of HyIR, defined by HydroBASINS Pfafstetter Level 1 boundaries (Lehner and Grill, 2013). Computation is performed independently and in parallel for each domain.

2.3 Climate forcing and water balance

2.3.1 CMIP6 forcing

HyIR is forced by monthly outputs from CMIP6 GCMs (Eyring et al., 2016), accessed through the Earth System Grid Federation (ESGF) (ESGF, 2026). Four variables are required for each HyIR run: precipitation (P), real evapotranspiration (RET), surface runoff (R_s), and total runoff (R). These fluxes, provided in units of $\text{kg m}^{-2} \text{s}^{-1}$, are converted to volumetric rates ($\text{m}^3 \text{s}^{-1}$) at the HydroUnit scale:

$$Q(t) = \frac{F(t) \cdot A_{\text{HU}}}{\rho} \quad (1)$$

where F is the CMIP6 flux, A_{HU} is the continental land area of the HydroUnit (excluding ocean and sea), computed from the grid cell geometry in an equal-area projection (EPSG:6933), and $\rho = 999.5 \text{ kg m}^{-3}$ is the density of water at approximately 12°C . Simulations cover the historical period (1850–2014) and future projections (2015–2100) under SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, depending on GCM availability.

2.3.2 Water balance partitioning

Effective precipitation P_{eff} ($\text{m}^3 \text{s}^{-1}$) is the net water input to the land surface after evapotranspiration, constrained to be non-negative:

$$P_{\text{eff}}(t) = \max(0, P(t) - RET(t)) \quad (2)$$



Months where evapotranspiration exceeds precipitation, which occur in arid regions and during dry seasons, yield $P_{\text{eff}} = 0$; the soil moisture deficit implied by such conditions is not tracked by HyIR, as land surface processes are delegated to the driving

140 GCM.

Total runoff from the GCM is partitioned into a surface component $\phi_s = R_s$ and a subsurface baseflow component ϕ_b :

$$\phi_b(t) = R(t) - R_s(t) \quad (3)$$

Both R and R_s are set to zero where negative values occur in the CMIP6 fields, and ϕ_s is additionally capped at P_{eff} to prevent surface runoff from exceeding the available water input.

145 Infiltration ϕ_{inf} is the residual of effective precipitation after surface runoff:

$$\phi_{\text{inf}}(t) = \max(0, P_{\text{eff}}(t) - \phi_s(t)) \quad (4)$$

The reconstructed infiltration flux ϕ_{inf} is used exclusively for storage accounting within the HyIR aggregation framework and should not be interpreted as an independent simulation of groundwater recharge or subsurface flow generation. Groundwater discharge to rivers remains inherited from the driving GCM through the baseflow component $\phi_b = R - R_s$ (Eq. (3)), preserving

150 the original climate–hydrology partitioning.

The HydroUnit water balance closes over all storage compartments and river inflow:

$$P_{\text{eff}}(t) = \Delta V_{\text{SWB}}(t) + \Delta V_{\text{aq}}(t) + Q_{\text{riv},\text{in}}(t) \quad (5)$$

where ΔV_{SWB} is the total storage change across all surface water bodies within the HydroUnit (Sect. 2.5), ΔV_{aq} is the groundwater storage variation (Sect. 2.6), and $Q_{\text{riv},\text{in}}$ ($\text{m}^3 \text{s}^{-1}$) is the total lateral inflow generated by this HydroUnit and routed to its river reaches (Sect. 2.7), including surface runoff, baseflow, and surface water body drainage. No soil moisture bucket or snow module is included; the full land surface partitioning is inherited from the GCM.

155

2.4 Climate zone classification

Hydrological processes, and in particular the direction of water exchange between surface water bodies and groundwater, differ substantially between arid and humid climates (Döll et al., 2003; Taylor et al., 2013). A binary climate classification of each HydroUnit into arid or humid regimes is therefore adopted to govern lake–groundwater exchange fluxes and baseflow routing (Sects. 2.5 and 2.6). The classification is based on the aridity index (AI), defined as the ratio of mean annual precipitation to potential evapotranspiration.

160

For the historical period, AI values are extracted from the historical climatology (1970–2000) of the Future Global Aridity Index and PET Database (Zomer and Trabucco, 2025) and assigned as a fixed per-HydroUnit classification for the entire historical simulation. For future projections, GCM- and SSP-specific AI fields from the same dataset are used, with the classification updated at the available projection time horizons; where a GCM-specific field is unavailable, a multi-model ensemble AI field is substituted for the corresponding SSP. HydroUnits with AI below a configurable threshold (default $\text{AI}_{\text{threshold}} = 0.65$) are classified as arid; all others as humid. This threshold follows the UNEP aridity convention (United Nations Environment Programme (UNEP), 1992) and is consistent with values used in global water resource assessments.

165



170 2.5 Surface water body storage

2.5.1 Input data and spatial intersection

Lake and reservoir storage is represented using Level 1 and Level 2 data from the Global Lakes and Wetlands Database (GLWD-1 and GLWD-2; Lehner and Döll 2004). GLWD-1 provides shoreline polygons for the 3 067 largest lakes (area $\geq 50 \text{ km}^2$) and 654 major reservoirs (storage capacity $\geq 0.5 \text{ km}^3$). GLWD-2 contains approximately 250 000 smaller water bodies with areas $\geq 0.1 \text{ km}^2$. Lake and reservoir areas are computed directly from the polygon geometries in an equal-area projection to ensure consistency with other spatial operations. Water bodies whose intersection area with a HydroUnit falls below a configurable threshold ($A_{\text{intersect}} < \theta \cdot A_{\text{min}}$, default $\theta = 10\%$) are excluded, where A_{min} is the area of the smallest water body in the dataset.

Maximum storage capacity $V_{\text{swb-max}}^i$ is taken from the GLWD VOLUME_CKM attribute where available. For water bodies without a volume attribute, capacity is estimated as $A_{\text{swb}}^i \times \bar{d}$, where $\bar{d} = 3.9 \text{ m}$ is the global mean lake depth (Messenger et al., 2016). Initial storage volumes at the start of each HyIR run are set to a configurable fraction of maximum capacity ($V_{\text{swb-init}}^i = \theta_{\text{init}} \cdot V_{\text{swb-max}}^i$, default $\theta_{\text{init}} = 40\%$).

2.5.2 Governing equations

The storage variation ΔV_{swb}^i of surface water body i within a HydroUnit is governed by:

$$\frac{\Delta V_{\text{swb}}^i(t)}{\Delta t} = P_{\text{eff} \rightarrow \text{swb}}^i(t) + \phi_{s \rightarrow \text{swb}}^i(t) - \phi_{\text{swb} \rightarrow \text{aq}}^i(t) - \phi_{\text{swb} \rightarrow \text{riv}}^i(t) \quad (6)$$

185 Effective precipitation falling directly on the water body surface is apportioned by area fraction:

$$P_{\text{eff} \rightarrow \text{swb}}^i(t) = P_{\text{eff}}(t) \cdot \frac{A_{\text{swb}}^i}{A_{\text{HU}}} \quad (7)$$

Surface runoff contribution is distributed proportionally to water body area within the HydroUnit:

$$\phi_{s \rightarrow \text{swb}}^i(t) = \phi_s(t) \cdot \frac{A_{\text{swb}}^i}{A_{\text{HU}}} \quad (8)$$

The exchange between surface water body i and the groundwater storage compartment is parameterized as a large-scale partitioning flux whose purpose is storage accounting and first-order separation of surface and subsurface compartments (Sect. 2.4); it is not intended to represent Darcy-scale groundwater dynamics or site-specific lake-groundwater interactions. In arid regions, the water body loses water to the groundwater compartment through a downward exchange flux:

$$\begin{aligned} \phi_{\text{swb} \rightarrow \text{aq}}^i(t) &= A_{\text{swb}}^i \cdot K_{\text{sat}}^i \cdot \omega_{\text{sat}}^i, \\ 0 &< \phi_{\text{swb} \rightarrow \text{aq}}^i(t) < \phi_{\text{inf}}^i(t) \end{aligned} \quad (9)$$



In humid regions, a compensating upward exchange flux is used to account for groundwater contributions to surface water
195 bodies:

$$\begin{aligned}\phi_{\text{swb} \rightarrow \text{aq}}^i(t) &= -A_{\text{swb}}^i \cdot K_{\text{sat}}^i \cdot \nabla h^i, \\ |\phi_{\text{swb} \rightarrow \text{aq}}^i(t)| &< \phi_b(t)\end{aligned}\quad (10)$$

where K_{sat}^i (m s^{-1}) is an effective exchange conductance parameter beneath the water body footprint, ω_{sat}^i is a saturation weighting factor (dimensionless), and $\nabla h^i = 10^{-3}$ is a prescribed large-scale exchange gradient used for consistency with the chosen parameterization (Döll et al., 2003).

200 Drainage from water body i to the river network follows a first-order outflow law scaled by a seasonal weighting factor:

$$\phi_{\text{swb} \rightarrow \text{riv}}^i(t) = C^i \cdot \delta^i(t) \cdot V_{\text{swb}}^i(t-1) \quad (11)$$

where C^i (s^{-1}) is an annual drainage coefficient and $\delta^i(t)$ is a dimensionless seasonal modulation factor. Storage at time step t is updated from the previous step and capped at maximum capacity:

$$V_{\text{swb}}^i(t) = \min(V_{\text{swb}}^i(t-1) + \Delta V_{\text{swb}}^i(t), V_{\text{swb-max}}^i) \quad (12)$$

205 When the updated volume exceeds $V_{\text{swb-max}}^i$, the surplus is routed as overflow to the river network at the same time step, preserving mass conservation. The total surface water body storage change at HydroUnit scale is:

$$\Delta V_{\text{SWB}}(t) = \sum_{i=1}^n \Delta V_{\text{swb}}^i(t) \quad (13)$$

2.6 Groundwater storage

2.6.1 Groundwater storage unit delineation

210 Representing groundwater storage at the global scale is constrained by the absence of a globally consistent dataset describing subsurface geometry, hydraulic properties, and regional flow systems. Available resources such as WHYMAP (Richs et al., 2011) and global hydrogeological simulations (De Graaf et al., 2015) offer useful conceptual overviews but do not provide a spatial representation readily compatible with the native GCM grid used in HyIR.

HyIR therefore adopts a pragmatic delineation of groundwater storage units based on the global depth-to-bedrock dataset of
215 Shangguan et al. (2017), which provides a spatially continuous estimate of soil depth at 10 km resolution. Two depth classes are defined from this dataset:

- **Shallow storage units:** depth between 5 and 20 m;
- **Deep storage units:** depth greater than 20 m.

220 Pixels where depth to bedrock is less than 5 m are excluded from the groundwater storage compartment, as such shallow depths are assumed insufficient for the development of a significant groundwater storage system. An 8-neighbor connectivity algorithm



groups contiguous pixels into coherent storage units, which are then polygonized into individual entities. The resulting global delineation of shallow and deep storage units is shown in Fig. D1. Storage capacity $V_{\text{aq-max}}$ of each unit is computed from pixel surface area, effective thickness and porosity.

The 5 m soil exclusion, the 20 m depth threshold, and the storage parameterizations introduced below are pragmatic first-order assumptions required to derive globally consistent storage-based groundwater indicators from currently available datasets. They should not be interpreted as physically universal hydrogeological parameters or as a substitute for regional groundwater flow modelling.

Effective porosity is assigned as a function of depth, reflecting the well-established decrease in pore space with increasing lithostatic pressure and compaction (De Graaf et al., 2015). For shallow storage units (5–20 m) a uniform porosity of 10% is applied. For deep storage units (depth $d > 20$ m), porosity decreases linearly from 10% at 20 m to 1% at 180 m, and continues to decrease from 1% to 0.1% beyond 180 m, representing the strongly reduced pore space at large depths where compaction and cementation progressively close fractures and intergranular voids. Spatially variable porosity values from the GLHYMPS dataset (Gleeson, 2018) may optionally be substituted to improve regional accuracy.

Initial groundwater storage volumes are set to:

$$V_{\text{aq,init}} = (d_{\text{avg}} - 1 \text{ m}) \cdot A_{\text{aq}} \cdot n_e \quad (14)$$

where d_{avg} is the area-weighted mean depth-to-bedrock within the storage unit, A_{aq} is the unit area, and n_e is the effective porosity. This initialization corresponds to a water table 1 m below the surface and is used as a simple, globally applicable starting condition for storage accounting; it is not intended to represent local hydrogeological equilibrium conditions. The sensitivity to initial aquifer storage, conditions, and the potential benefit of a spin-up period, are left for future investigation.

2.6.2 Governing equations

The net variation in groundwater storage volume over a time step is:

$$\frac{\Delta V_{\text{aq}}(t)}{\Delta t} = \phi_{\text{inf}}(t) - \phi_{\text{aq} \rightarrow \text{riv}}(t) + \sum_{i=1}^n \phi_{\text{swb} \rightarrow \text{aq}}^i(t) \quad (15)$$

Baseflow discharge to the river network depends on the climatic regime. In HyIR, the groundwater storage compartment does not represent regional groundwater connectivity, transmissivity, or travel-time effects; instead, baseflow is inherited from the GCM as $\phi_b = R - R_s$ and allocated to rivers or surface water bodies depending on the aridity classification. In arid regions, all baseflow is directed to rivers:

$$\phi_{\text{aq} \rightarrow \text{riv}}(t) = \phi_b(t) \quad (16)$$



In humid regions, part of the baseflow is accounted for by groundwater exfiltration into surface water bodies, and the river contribution is:

$$\phi_{\text{aq} \rightarrow \text{riv}}(t) = \phi_b(t) + \sum_{i=1}^n \phi_{\text{swb} \rightarrow \text{aq}}^i(t),$$

$$\phi_{\text{swb} \rightarrow \text{aq}}^i(t) < 0$$
(17)

The condition $\phi_{\text{swb} \rightarrow \text{aq}}^i(t) < 0$ selects only exfiltration fluxes (inflow to the lake). Storage volume is updated and bounded:

$$V_{\text{aq}}(t) = \min(V_{\text{aq}}(t-1) + \Delta V_{\text{aq}}(t), V_{\text{aq-max}})$$
(18)

Storage is constrained to remain non-negative. Any excess above $V_{\text{aq-max}}$ is transferred to surface runoff at the same time step. For HydroUnits with no intersecting storage unit, infiltration is routed directly to the river network.

2.7 River network routing

2.7.1 Network data and reach upscaling

River routing in HyIR uses the HydroRIVERS and HydroBASINS datasets (Lehner and Grill, 2013), derived from the HydroSHEDS framework (Lehner et al., 2008) based on the 15 arc-second SRTM digital elevation model. HydroRIVERS provides a globally consistent vector representation of the river network segmented into over 8 million individual reaches, each carrying attributes including length, upstream contributing area, long-term mean discharge, Strahler stream order, and upstream–downstream connectivity.

River routing can be performed on the full HydroRIVERS network and can in principle be run at the native reach resolution of over 8 million individual reaches. At global scale this level of detail is computationally prohibitive and provides little benefit for large-scale flow representation. HyIR therefore includes a river upscaling module that aggregates the original network to a coarser Pfafstetter level prior to routing. The default configuration uses Pfafstetter Level 6 of HydroBASINS (average basin area $\approx 3000 \text{ km}^2$), reducing the routing problem from millions of reaches to a tractable number of representative reaches, one per Level 6 basin. The aggregation algorithm selects the main stem of each basin based on Strahler stream order and accumulates hydraulic and geometric attributes from the constituent reaches, including upstream contributing area, reach length, bed slope, and distance to the outlet, into a single representative reach. Higher Pfafstetter levels can be selected to increase routing resolution at the cost of greater computational demand.

2.7.2 Lateral inflow assembly

The total volumetric inflow to the river network from HydroUnit H is assembled from three contributions:

$$\frac{\Delta V_{\text{riv}}(t)}{\Delta t} = \phi_{s \rightarrow \text{riv}}(t) + \phi_{\text{aq} \rightarrow \text{riv}}(t) + \sum_{i=1}^n \phi_{\text{swb} \rightarrow \text{riv}}^i(t)$$
(19)

Surface runoff reaching the river is reduced by the fraction intercepted by surface water bodies within the HydroUnit:

$$\phi_{s \rightarrow \text{riv}}(t) = \phi_s(t) \cdot (1 - f_{\text{swb}})$$
(20)



where f_{swb} is the fractional area coverage of surface water bodies within the HydroUnit.

This monthly inflow is then disaggregated to individual river reaches proportionally by elementary catchment area:

$$F^i(t) = \frac{\Delta V_{\text{riv}}^H(t)}{\Delta t} \cdot \frac{A^i}{\sum_{j \in H} A^j} \quad (21)$$

where A^i is the area of the elementary catchment associated with reach i within HydroUnit H .

280 2.7.3 Muskingum routing

River discharge is computed at a daily time step for each reach i using a semi-implicit Muskingum formulation (Cunge, 1969; David et al., 2013; Flipo et al., 2023). The Muskingum storage equation relates reach volume to inflow and outflow:

$$V^i(t) = k^i [\alpha^i (Q_{\text{in}}^i(t) + F^i(t)) + (1 - \alpha^i) Q_{\text{out}}^i(t)] \quad (22)$$

where k^i (s) is the reach-specific travel time constant, α^i is a dimensionless weighting parameter, Q_{in}^i is the upstream inflow, and F^i is the lateral inflow.

Applying a semi-implicit temporal discretization with weighting factor $\theta \in [0, 1]$:

$$Q_{\text{out}}^i(t+1) = A^i (Q_{\text{in}}^i(t+1) + F^i(t+1)) + B^i (Q_{\text{in}}^i(t) + F^i(t)) + C^i Q_{\text{out}}^i(t) \quad (23)$$

The Muskingum coefficients are:

$$A^i = \frac{\theta - \frac{\alpha^i k^i}{\Delta t}}{\theta + \frac{(1 - \alpha^i) k^i}{\Delta t}} \quad (24)$$

$$B^i = \frac{(1 - \theta) + \frac{\alpha^i k^i}{\Delta t}}{\theta + \frac{(1 - \alpha^i) k^i}{\Delta t}} \quad (25)$$

$$C^i = \frac{\frac{(1 - \alpha^i) k^i}{\Delta t} - (1 - \theta)}{\theta + \frac{(1 - \alpha^i) k^i}{\Delta t}} \quad (26)$$

All three coefficients are constrained to remain non-negative. The condition $A^i + B^i + C^i = 1$ is enforced, guaranteeing mass conservation. Where the stability bounds on k^i are violated, k^i is clipped to the nearest stable value before renormalization. The weighting parameter α and the numerical scheme factor θ are configurable; default values are $\alpha = 0.25$, consistent with the range recommended for natural stream channels (Hamed et al., 2016), and $\theta = 0.5$ (Crank–Nicolson discretization).



2.7.4 Matrix formulation

At each daily time step, the full river network within each continental domain is solved simultaneously as a sparse linear system. The discharge vector $\mathbf{Q}$ collects the outflows from all N_R reaches. Network topology is encoded in the connectivity matrix $\mathbf{N}$:

$$N_{ij} = \begin{cases} 1 & \text{if reach } j \text{ flows into reach } i \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

The full system at time step $t + 1$ is:

$$(\mathbf{I} - \mathbf{AN}) \mathbf{Q}^{t+1} = \mathbf{AF}^{t+1} + \mathbf{B}(\mathbf{NQ}^t + \mathbf{F}^t) + \mathbf{CQ}^t \quad (28)$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are diagonal matrices containing the per-reach Muskingum coefficients. This formulation eliminates sequential reach-by-reach processing and allows the entire river network of a continental domain to be advanced in one sparse matrix solve per time step.

2.8 Parameterization

2.8.1 Muskingum travel time parameter k

The reach-specific travel time constant k is estimated from the time of concentration T_c of the river system, computed as the geometric mean of two empirical estimates, Pasini (Pasini, 1914) and Ventura (de Almeida et al., 2014):

$$T_c = \sqrt{T_{c, \text{Pasini}} \cdot T_{c, \text{Ventura}}} \quad (29)$$

where:

$$T_{c, \text{Pasini}} = F_p (L_{\text{riv}} A_{\text{bv}})^{1/3} I^{-0.5} \quad (30)$$

$$T_{c, \text{Ventura}} = F_v A_{\text{bv}}^{0.5} I^{-0.5} \quad (31)$$

with $F_p = 3888 \text{ s km}^{-1}$ and $F_v = 4578 \text{ s km}^{-1}$ the Pasini and Ventura coefficients respectively, L_{riv} (km) the cumulative river length of the upstream watersheds, A_{bv} (km^2) the upstream watershed area, and I (m m^{-1}) the mean watershed slope.

T_c represents the total travel time from headwater to the basin outlet. It is distributed spatially across individual reaches using the relative transfer time index I_{tr} (Flipo et al., 2012; Korkmaz, 2007):

$$I_{\text{tr}}(i) = \frac{dl_i}{|z(i) - z_{ds}(i)|^{0.5}} \cdot \frac{1}{A_{\text{up}}(i)^\beta} \quad (32)$$

where dl_i (m) is the reach length, $z(i)$ and $z_{ds}(i)$ (m) are the elevations of reach i and its immediate downstream neighbor respectively, $A_{\text{up}}(i)$ (m^2) is the upstream drainage area, and $\beta = 0.25$ (Korkmaz, 2007). The cumulative index from reach i to the basin outlet is:

$$I_{\text{tr}, i \rightarrow \text{outlet}} = \sum_{k=i}^{\text{outlet}} I_{\text{tr}}(k) \quad (33)$$



The distributed Muskingum k for each reach is:

$$k(i) = \frac{I_{tr, i \rightarrow \text{outlet}} - I_{tr, ds \rightarrow \text{outlet}}}{I_{tr, \max}} \cdot T_c \quad (34)$$

where ds denotes the immediate downstream reach and $I_{tr, \max}$ is the sum of all I_{tr} values in the basin. This formulation ensures that reaches with larger drainage areas and gentler slopes receive proportionally larger travel time contributions.

2.8.2 Hydraulic conductivity and soil porosity

Saturated hydraulic conductivity K_{sat} (m s^{-1}) at lake and reservoir footprints is derived from the global measured and modeled dataset of Surya et al. (2020). Spatial gaps are filled using lithological classifications from the Global Lithological Map (GLiM; Hartmann and Moosdorf 2012), with per-lithology median K_{sat} values interpolated from the Surya et al. database. Soil porosity at saturation ω_{sat} is estimated as:

$$\omega_{\text{sat}} = 1 - \frac{\rho_b}{\rho_p} \quad (35)$$

where bulk density ρ_b is obtained from the SoilGrids global database (ISRIC – World Soil Information, 2020) as a depth-integrated average over the 0–200 cm profile, and particle density $\rho_p = 2.65 \text{ g cm}^{-3}$ is assumed for mineral soils (Hillel, 1998).

2.8.3 Drainage coefficient and seasonal weighting

The drainage coefficient C^i (s^{-1}) governs the rate at which a surface water body drains to the river network (Eq. (11)), and represents the maximum fractional volume discharged per unit time. It is derived from an exponential decay assumption: imposing that 99% of an initial volume drains within 6 months gives:

$$C^i = -\frac{\ln(0.01)}{6 \times 30 \times 86400} \approx 2.96 \times 10^{-7} \text{ s}^{-1} \quad (36)$$

The seasonal modulation factor $\delta^i(t)$ varies sinusoidally with the annual cycle, with a hemisphere-aware phase:

$$\delta^i(t) = \bar{\delta} + A_\delta \cdot \sin\left(\frac{2\pi(t + \varphi)}{12}\right) \quad (37)$$

where $\bar{\delta} = 1/12$ ensures that the annual mean drainage sums to unity, $A_\delta = 0.02$ is the seasonal amplitude, and φ is a hemisphere-aware phase shift such that drainage peaks during the high-flow season: boreal summer (July–August) in the Northern Hemisphere, austral summer (January–February) in the Southern Hemisphere, and with a reduced amplitude for equatorial units where seasonality is weak.

2.9 Calibration parameters

Although the HyIR hydrological aggregator is designed to preserve the climate signal and physical consistency of the driving GCM, a small set of three scalar correction parameters is provided to address known systematic limitations of coarse-resolution GCM outputs and global empirical parameterizations. These corrections are applied within the HyIR aggregation layer; they



do not aim to recalibrate the framework against observations or to alter the GCM forcing fields themselves. Instead, they apply targeted, large-scale adjustments to specific representations (e.g., partitioning and routing) where structural biases are well documented in the literature, allowing a more robust translation of climate projections into water availability indicators without compromising scenario consistency.

The **time-of-concentration corrector** $\gamma_{T_c} \in (0, 1]$ rescales the travel time estimated by the geometric mean of the Pasini and Ventura formulas (Eq.(29)):

$$T_c^* = \gamma_{T_c} \cdot T_c \quad (38)$$

Empirical T_c estimates are known to overestimate travel times, particularly in large flat basins (Grimaldi et al., 2012). Reducing T_c shortens reach-level k values uniformly, accelerating wave propagation and improving the timing of discharge peaks.

The **infiltration-to-runoff redirector** $\gamma_{\text{inf}} \in (-1, 1)$ redistributes a fraction of the partitioning between surface runoff and infiltration while conserving their sum.

For $\gamma_{\text{inf}} > 0$, a fraction of infiltration is redirected to surface runoff:

$$\phi_s^*(t) = \phi_s(t) + \gamma_{\text{inf}} \cdot \phi_{\text{inf}}(t) \quad (39)$$

$$\phi_{\text{inf}}^*(t) = (1 - \gamma_{\text{inf}}) \cdot \phi_{\text{inf}}(t) \quad (40)$$

For $\gamma_{\text{inf}} < 0$, a fraction of surface runoff is redirected to infiltration:

$$\phi_{\text{inf}}^*(t) = \phi_{\text{inf}}(t) + |\gamma_{\text{inf}}| \cdot \phi_s(t) \quad (41)$$

$$\phi_s^*(t) = (1 - |\gamma_{\text{inf}}|) \cdot \phi_s(t) \quad (42)$$

In both cases the redirected flux is capped so that neither component becomes negative, preserving mass balance. This bidirectional correction addresses well-documented uncertainties in the simulation of runoff by coarse-resolution GCMs (Hou et al., 2023).

The **baseflow seasonal corrector** $\gamma_b \in \mathbb{R}$ is a multiplicative adjustment applied to ϕ_b , with a hemisphere-aware seasonal profile:

$$\phi_b^*(t) = \phi_b(t) \cdot \begin{cases} 1 + w(t) \gamma_b, & \gamma_b > 0 \\ [1 + (1 - w(t)) |\gamma_b|]^{-1}, & \gamma_b < 0 \end{cases} \quad (43)$$

where $w(t) \in [0, 1]$ is the seasonal weight, maximal during the high-flow season of each hemisphere. For $\gamma_b > 0$, baseflow is amplified, with the largest increase during the high-flow season; for $\gamma_b < 0$, baseflow is dampened, with the largest reduction during the low-flow season, so that the seasonal contrast is preserved in both directions. This correction addresses uncertainties in the seasonal dynamics of GCM-derived subsurface runoff (Hou et al., 2023). Mass balance is preserved, as the adjusted baseflow is drawn directly from aquifer storage (Sect. 2.6).

All three parameters default to their neutral values: $\gamma_{T_c} = 1$, $\gamma_{\text{inf}} = 0$, $\gamma_b = 0$. The values selected for the TaiESM1 application presented in this paper, and their sensitivity, are discussed in Sect. 4.



380 3 Model evaluation framework

3.1 River discharge datasets

3.1.1 GRDC discharge records

Daily river discharge records from The Global Runoff Data Centre (GRDC) (2025) are used as the primary evaluation dataset. Stations are spatially matched to HyIR upscaled river reaches following the methodology described in Sect. 3.2, retaining only
385 stations on large river basins. Only data prior to 2015 are considered to remain within the CMIP6 historical period. Stations with fewer than 15 years of records meeting a configurable minimum daily coverage (default: 30%) per year are excluded. Each retained station is evaluated over its full temporal overlap with the simulation period.

An outlier treatment is applied per station to the retained discharge records, consisting of a log-transform followed by a robust z-score screening based on the median and median absolute deviation (MAD), with a permissive threshold of $|z_{\text{robust}}| > 3$ to
390 remove only extreme erroneous values while preserving natural discharge variability.

3.1.2 GloFAS reanalysis dataset

GRDC station coverage is geographically uneven. Large river basins in Africa and Asia are particularly underrepresented: matched GRDC stations cover basins representing approximately 4.2% and 5.5% of the continental land area of Africa and Asia respectively, a coverage insufficient for continental-scale validation. To extend spatial coverage in these regions, the
395 Global Flood Awareness System (GloFAS; Harrigan et al., 2020) reanalysis discharge product (ERA5-forced, 1979–2014) is used as a complementary dataset where GRDC records are absent.

GloFAS provides daily gridded discharge estimates at $0.1^\circ \times 0.1^\circ$ resolution derived from a physically based land surface and routing model forced with ERA5 reanalysis. GloFAS pixels are matched to HyIR river reaches following the same outlet-priority logic as GRDC stations (Sect. 3.2), retaining only pixels with upstream area above the same threshold and within the
400 same area-consistency tolerance. GloFAS is treated as an independent benchmark and results from GloFAS-validated reaches are reported separately from GRDC-validated reaches.

3.2 Station and pixel matching methodology

GRDC stations and GloFAS pixels are matched to HyIR upscaled river reaches using a two-priority spatial algorithm applied independently for each large river system. A river system is defined as large if its outlet reach has an upstream contributing
405 area exceeding $50\,000\text{ km}^2$.

For each large river outlet, the algorithm first searches for an observation point within a configurable buffer of the outlet pour point (default: 5 km). If no match is found, the search is extended to the main-stem backbone of the river within a wider configurable buffer (default: 10 km), retaining the candidate closest to the outlet. A single observation point is retained per outlet.



410 A drainage area consistency check discards matches where the relative difference between the observation catchment area and the HyIR upstream area exceeds a configurable tolerance (default: 30%), removing cases where the observation point drains a substantially different area from the modeled reach. All buffer distances and data availability criteria are configurable.

3.3 Performance metrics

3.3.1 Kling–Gupta Efficiency

415 Model performance is assessed using the Kling–Gupta Efficiency (KGE; Gupta et al., 2009):

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (44)$$

where r is the Pearson correlation coefficient between HyIR-derived and observed daily discharge, $\alpha = \sigma_s/\sigma_o$ is the ratio of standard deviations, and $\beta = \mu_s/\mu_o$ is the ratio of means. KGE ranges from $-\infty$ to 1, with $\text{KGE} = 1$ indicating perfect agreement. Values above -0.41 indicate that the framework outperforms a benchmark that predicts a constant value equal to
420 the long-term mean observed discharge over the whole observation period, which yields $\text{KGE} = 1 - \sqrt{2} \approx -0.41$ (Knoben et al., 2019). Values below this threshold indicate that the simulation performs worse than this trivial predictor. It should be noted that for drought-oriented assessment, a constant mean-flow predictor is arguably a weak benchmark, since it carries no information on low-flow dynamics; the threshold therefore represents a conservative lower bound on acceptable performance in the context of the present application. KGE is computed over all days with valid observed and HyIR-derived discharge in
425 the overlap period.

3.3.2 Drought-focused KGE

Standard KGE weighs variability through the ratio of standard deviations α , which is dominated by high-flow periods. To evaluate framework performance on low-flow conditions, which are particularly relevant for industrial water availability assessment, a drought-focused variant is proposed by replacing α with the ratio of QMNA5 values:

$$\text{KGE}_{\text{drought}} = 1 - \sqrt{(r-1)^2 + (\text{QMNA5}_{\text{ratio}}-1)^2 + (\beta-1)^2} \quad (45)$$

where $\text{QMNA5}_{\text{ratio}} = \text{QMNA5}_s/\text{QMNA5}_o$, with subscripts s and o denoting simulated and observed values respectively. QMNA5 is the 5-year return period of the annual minimum monthly mean discharge, computed as the 20th percentile of annual minima (Tallaksen and van Lanen, 2004), and is a standard hydrological indicator of chronic low-flow conditions. Its ratio quantifies the framework's capacity to reproduce the magnitude of multi-annual drought-level flows, replacing the
435 variability term α with a low-flow-specific measure of bias.

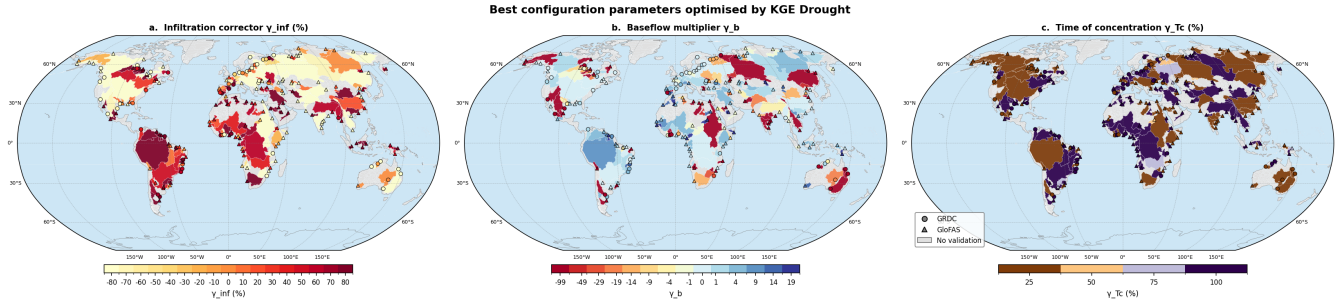


Figure 3. Spatial distribution of the best-performing calibration parameter combination per river basin, selected by maximizing $KGE_{drought}$. a: infiltration corrector γ_{inf} (%); b: baseflow seasonal multiplier γ_b ; c: time-of-concentration corrector γ_{Tc} (%). Basin colors indicate the selected parameter value. Circles and triangles show matched GRDC stations and GloFAS pixels respectively.

4 Application to TaiESM1

This section presents an application of HyIR forced with the TaiESM1 CMIP6 general circulation model as a proof of concept, illustrating the framework workflow and assessing the ability of the coupled climate–hydrology chain to match observed discharge. The validation results presented here reflect the joint performance of TaiESM1 and HyIR and should not be interpreted as an evaluation of HyIR in isolation.

4.1 GCM and aggregator assessment description

TaiESM1 (Taiwanese Earth System Model; Lee et al., 2020) is selected among the CMIP6 GCMs that provide all four variables required by the HyIR hydrological aggregator (P , RET , R , R_s) for both the historical period and future SSP scenarios. TaiESM1 operates at a native resolution of $1.25^\circ \times 0.9^\circ$ and provides historical simulations covering 1850–2014 (ensemble member r1i1p1f1). HyIR is applied on the native TaiESM1 grid, yielding approximately 16 000 continental HydroUnits (out of 55 000 global grid cells).

All framework parameters, including routing, hydrological, climate classification, and validation matching parameters, are set to their default values as described in Sects. 2.8, 2.4, and 3.2.

4.2 Calibration parameter selection

The three calibration parameters described in Sect. 2.9 (γ_{Tc} , γ_{inf} , γ_b) are tuned against GRDC and GloFAS observations using a discrete grid search over a predefined parameter space. The tested values span $\gamma_{inf} \in \{-0.8, -0.7, \dots, 0.7, 0.8\}$ in steps of 0.1, $\gamma_b \in \{-99, -49, -29, -19, -14, -9, -4, -1, 0, 1, 4, 9, 14, 19\}$, and $\gamma_{Tc} \in \{0.25, 0.5, 0.75, 1.0\}$, resulting in a total of 952 tested configurations per continental domain. Parameter selection is performed independently for each river basin, retaining the configuration maximizing $KGE_{drought}$. Basins with no observational match are assigned the best configuration of their spatially neighboring matched basins.

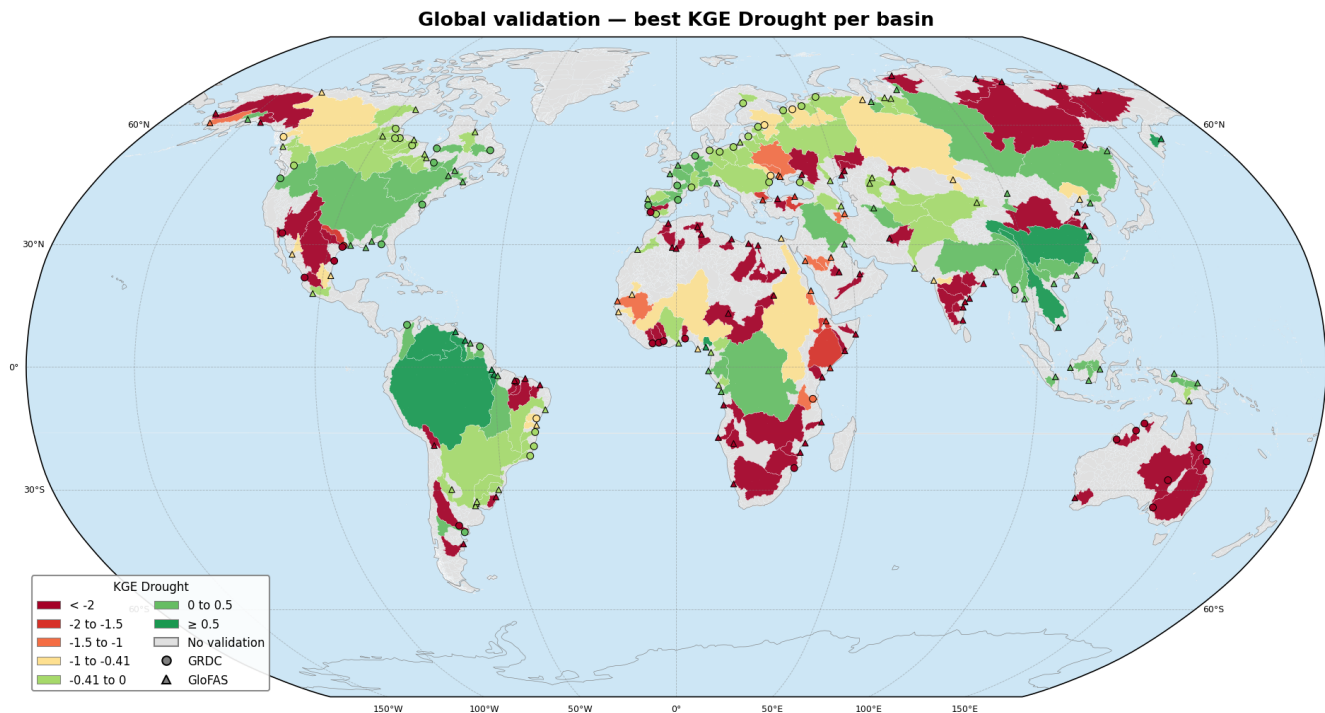


Figure 4. Spatial distribution of $KGE_{drought}$ per river basin for HyIR forced with TaiESM1 historical climate (1850–2014), at the best-performing calibration configuration. Basin colors indicate $KGE_{drought}$ category. Circles: matched GRDC stations; triangles: matched GloFAS pixels.

The spatial distribution of the selected parameter values is shown in Fig. 3. The infiltration corrector γ_{inf} tends toward higher values in tropical and monsoonal regions, suggesting a systematic tendency of TaiESM1 to partition an excess fraction of effective precipitation into subsurface fluxes in humid, high-runoff environments. The baseflow seasonal multiplier γ_b shows greater variability across regions, with higher values selected in mid-latitude basins where the seasonal signal of subsurface runoff is more pronounced. The time-of-concentration corrector γ_{T_c} is predominantly selected at values below unity across most continental domains, consistent with the known tendency of empirical T_c formulas to overestimate travel times in large flat basins.

4.3 Historical validation results

The spatial distribution of $KGE_{drought}$ across all validated basins is presented in Fig. 4. Performance varies substantially across regions, with the best results concentrated in South America and parts of Asia, and the weakest in Australia, Siberia, and Arctica. At the global scale, 59% of the validated area achieves $KGE_{drought} > -0.41$, with 40.4% scoring above zero and 13.1% reaching $KGE_{drought} \geq 0.5$.

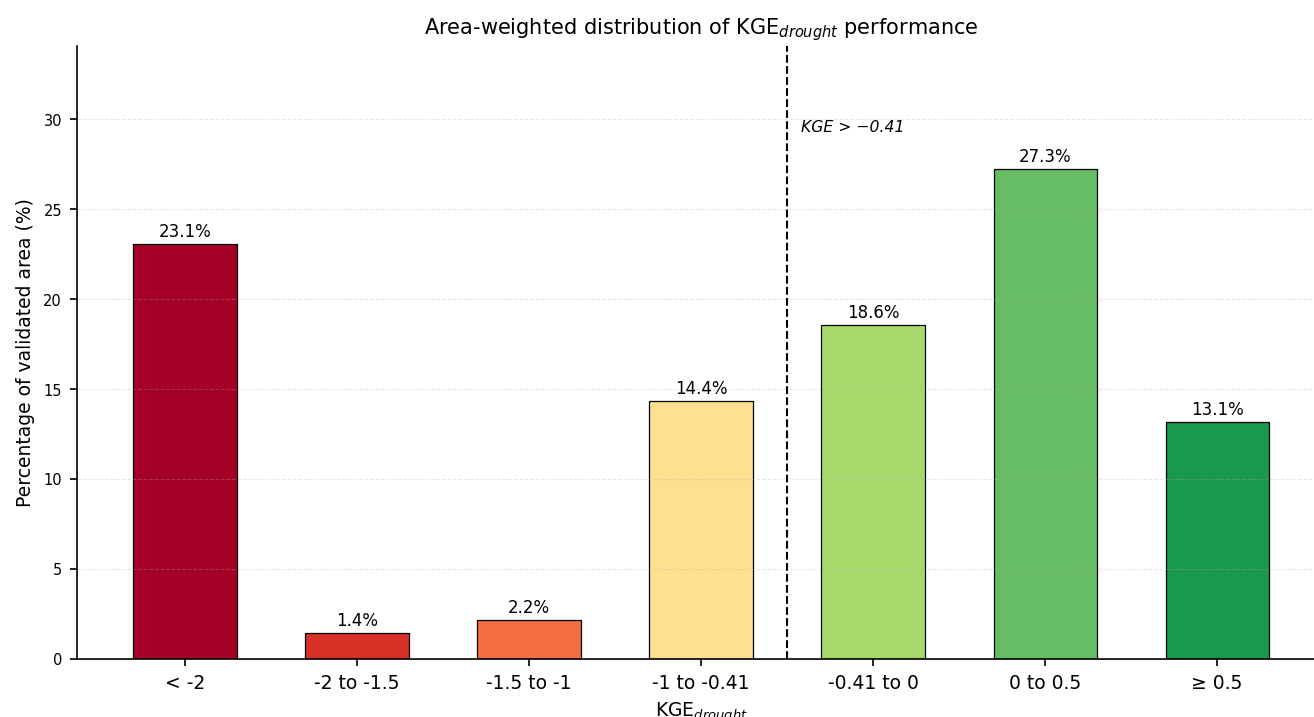


Figure 5. Area-weighted distribution of $KGE_{drought}$ performance across all validated basins globally. Bar heights represent the percentage of total validated area within each $KGE_{drought}$ category. The dashed line indicates the $KGE = -0.41$ acceptable performance threshold (Knoben et al., 2019).

The area-weighted global distribution of $KGE_{drought}$ (Fig. 5) and the regional breakdown (Table 1) confirm these patterns. Performance is strongest in South America, where 91.6% of the validated area exceeds the threshold of $KGE_{drought} = -0.41$, and in North America and Asia, where 82.9% and 82.5% of the validated area respectively score above the same threshold. The arctic, Australia and Siberia show the weakest results, with 90.6%, 77.8% and 69.3% of their validated areas respectively scoring below -0.41 .

To illustrate framework behavior at the river scale, Fig. 6 presents HyIR-derived and observed monthly mean discharge for a selection of major river basins spanning different continental regions, climates, and basin sizes.

475 5 Discussion

5.1 Scope and design

HyIR hydrological aggregator is designed as an aggregation framework that translates CMIP6 climate projections into scenario-consistent estimates of freshwater availability. In this positioning, HyIR deliberately delegates the climate–hydrology relation-



Table 1. Regional $KGE_{drought}$ classification: validated area (10^3 km^2) and percentage of validated area by performance category for each continental domain and globally.

Region	Validated area (10^3 km^2)	$KGE_{drought}$ category (10^3 km^2 , %)				
		< -1	$-1 \text{ to } -0.41$	$-0.41 \text{ to } 0$	$0 \text{ to } 0.5$	≥ 0.5
Africa	16 761	6940 (41.4%)	5146 (30.7%)	700 (4.2%)	3921 (23.4%)	53 (0.3%)
Asia	12 102	1791 (14.8%)	324 (2.7%)	2396 (19.8%)	4907 (40.5%)	2683 (22.2%)
Australia	2 243	1744 (77.8%)	0 (0.0%)	64 (2.8%)	435 (19.4%)	0 (0.0%)
Europe	8 444	2180 (25.8%)	695 (8.2%)	3832 (45.4%)	1737 (20.6%)	0 (0.0%)
N. America	9 644	1438 (14.9%)	210 (2.2%)	2423 (25.1%)	5574 (57.8%)	0 (0.0%)
S. America	13 424	1016 (7.6%)	109 (0.8%)	3991 (29.7%)	1250 (9.3%)	7059 (52.6%)
Siberia	9 162	3835 (41.9%)	2468 (26.9%)	262 (2.9%)	2544 (27.8%)	53 (0.6%)
Arctic	3 113	1021 (32.8%)	1796 (57.7%)	245 (7.9%)	50 (1.6%)	0 (0.0%)
Global	74 892	19966 (26.7%)	10749 (14.4%)	13912 (18.6%)	20417 (27.3%)	9848 (13.1%)

ship to the driving GCM and preserves the internal physical consistency of each realization. HyIR is therefore not intended to replace global hydrological models, but to provide a consistent translation of GCM outputs into water availability indicators with explicit storage accounting. HyIR inherits the scenario structure of the IPCC socioeconomic framework, producing hydrological projections that are fully consistent with the SSP forcing across the historical period and through 2100. This makes the framework applicable to the full CMIP6 ensemble, enabling assessments of freshwater availability under contrasting emission trajectories.

A key contribution of HyIR is the explicit inclusion of a groundwater compartment, with groundwater storage accounted for and distinguished from surface water resources. While this representation remains necessarily simplified at the planetary scale and should not be interpreted as a fully resolved hydrogeological model, it provides a pragmatic first-order estimate of subsurface water resources by delineating groundwater storage units from global depth-to-bedrock data. This explicit storage accounting helps address a common limitation of large-scale climate impact studies, which often emphasize river discharge and provide only limited information on groundwater stocks. Future developments could strengthen this component by integrating additional hydrogeological information (e.g., aquifer geometry, hydraulic properties, and regional groundwater connectivity).

This design positions HyIR as a tool for prospective resilience assessment of water-dependent industries under a changing climate. Mining is used here as a motivating application because operations often concentrate freshwater demands in arid and semi-arid regions that are exposed to projected increases in water stress under high-emission scenarios (Masson-Delmotte et al., 2021; Lee and Romero, 2023). More generally, the framework is applicable to any water-dependent infrastructure or industrial sector requiring long-term, scenario-consistent assessments of freshwater availability under climate change (e.g.,

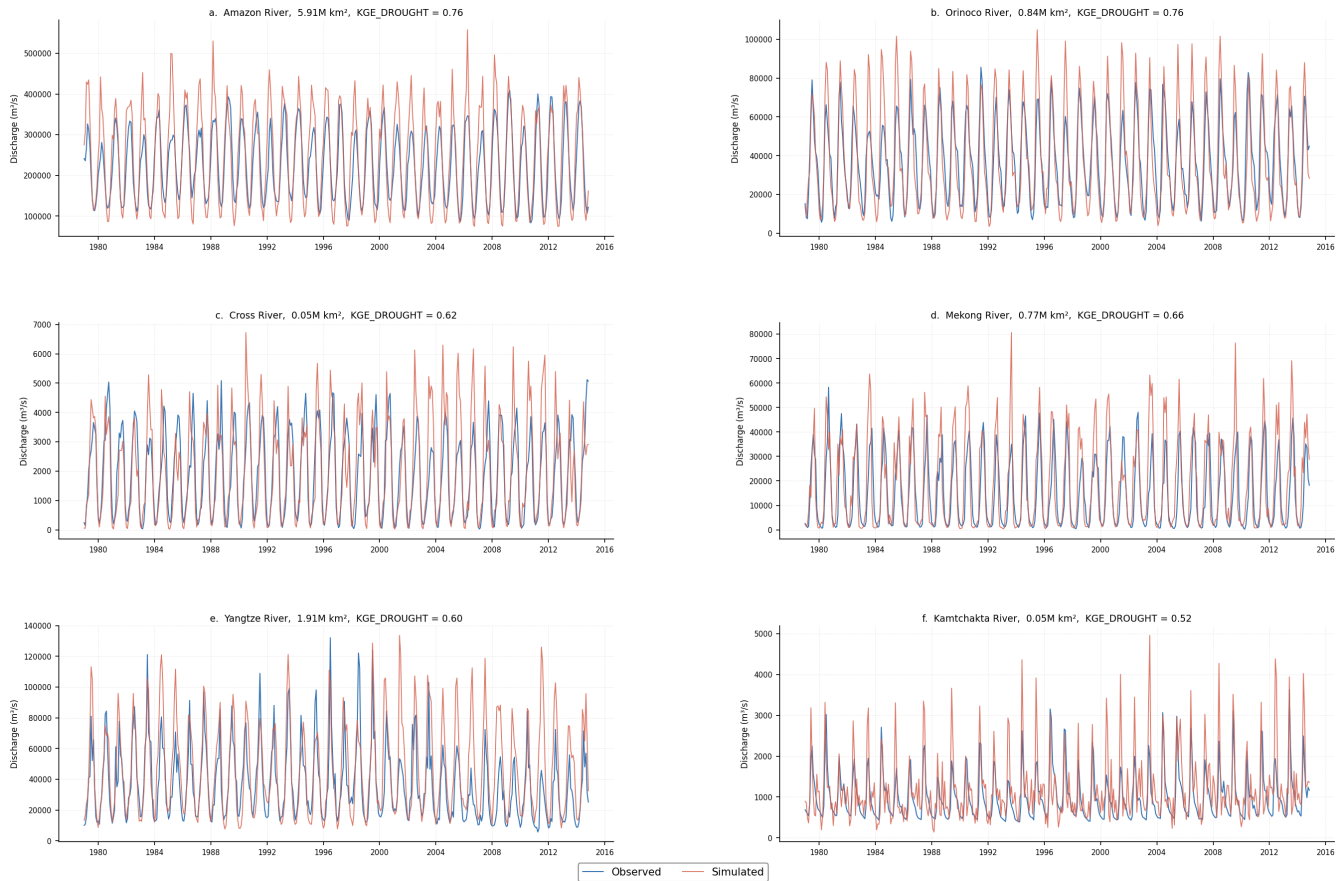


Figure 6. Simulated (orange) and observed (blue) monthly mean discharge ($\text{m}^3 \text{s}^{-1}$) at selected major river basin outlets, for HyIR forced with TaiESM1 historical climate.

energy production, agriculture, industrial manufacturing, urban water supply, hydrogen production, data centers, and critical infrastructure planning).

5.2 Calibration strategy and GCM biases

While the framework delegates the climate–hydrology relationship to the GCM, a calibration step is included to address known limitations and improve the interpretability of water availability estimates. The three calibration parameters introduced in Sect. 2.9 are pragmatic large-scale correction parameters targeting documented structural limitations of coarse-resolution GCM outputs. The infiltration-to-runoff redirector γ_{inf} addresses the tendency of coarse climate models to misrepresent the partitioning of effective precipitation between surface and subsurface pathways (Hou et al., 2023), a limitation that is particularly pronounced in humid tropical regions where infiltration-excess processes operate at sub-grid scales. The time-of-concentration corrector γ_{T_c} compensates for the systematic overestimation of travel times by the Pasini and Ventura empirical formulas,



which are known to perform poorly in large flat basins (Grimaldi et al., 2012). The baseflow seasonal corrector γ_b addresses uncertainties in the seasonal dynamics of GCM-derived subsurface runoff, which can misrepresent the amplitude of the annual baseflow cycle (Hou et al., 2023).

510 The spatially coherent patterns of selected parameter values visible in Fig. 3 reflect regionally consistent biases in the forcing's hydrological output. The bias in subsurface partitioning is climate-regime dependent, consistent with the expectation that land surface schemes in coarse GCMs perform differently across aridity gradients. In this sense, the calibration framework serves as both a performance-improvement step and a diagnostic tool for characterizing GCM-specific hydrological limitations at the large basin scale.

515 5.3 Limitations

Because HyIR is designed as an aggregation framework for translating CMIP6 projections into water availability indicators, several processes are represented in a simplified manner or deliberately omitted. The current version of HyIR does not represent snow accumulation and melt, canopy interception, or groundwater lateral flow. The absence of a snow module limits the applicability of the framework in snow-dominated basins, where the seasonal storage and release of precipitation as snowpack
520 constitutes the primary control on discharge timing. Canopy interception, which can account for a non-negligible fraction of gross precipitation in densely vegetated regions, is entirely delegated to the GCM land surface scheme. Groundwater is represented as a set of independent storage units with no lateral connectivity, which precludes the representation of inter-basin groundwater transfers and regional groundwater flow systems that control low-flow behavior in many arid settings. This simplification is adopted to retain global applicability and compatibility with CMIP6 forcing while keeping assumptions
525 transparent. Despite this, the explicit groundwater storage compartment provides a first-order differentiation between surface and subsurface freshwater resources. Future developments could incorporate additional hydrogeological information (e.g., regional connectivity and hydraulic properties) where datasets allow.

Human water uses are not accounted for in the current framework. Irrigation withdrawals, reservoir management, and groundwater pumping can substantially alter the natural discharge regime in heavily managed basins, introducing system-
530 atic biases between HyIR-derived discharge estimates and observed discharge that are unrelated to the framework's physical representation. This limitation is most consequential in large irrigated river systems where anthropogenic withdrawals represent a dominant term in the water balance. Incorporating human water use and demand will be a priority in future development of the HyIR framework.

5.4 Potential applications

535 Beyond the proof-of-concept presented here, HyIR is designed to be applied across the full CMIP6 GCM ensemble and SSP scenario space. Multi-model ensemble runs would allow uncertainty in hydrological projections to be partitioned between GCM structural differences and scenario uncertainty, providing a more complete picture of the range of possible freshwater futures. At the site level, the explicit representation of water availability by source type, distinguishing river discharge, lake and reservoir storage, and groundwater, makes HyIR directly applicable to water sourcing strategy assessment for industrial



540 facilities, allowing evaluation of the relative exposure of surface water and groundwater supplies to climate-driven stress under different SSPs. Combined with georeferenced datasets of water-consuming industrial sites, the framework can support prospective assessments of the long-term viability of water-intensive operations, informing adaptation planning at the global scale for sectors whose resilience depends on reliable freshwater access.

6 Conclusions

545 HyIR 0.1, a global-scale hydrological aggregation framework, designed to translate CMIP6 climate projections into scenario-consistent estimates of freshwater availability across rivers, lakes and reservoirs, and groundwater storage compartments, is presented here. The hydrological aggregator propagates GCM precipitation, evapotranspiration, and runoff fields through three coupled storage and routing components operating on the native GCM grid, covering the full CMIP6 temporal envelope from 1850 to 2100 under multiple SSPs.

550 Applied to TaiESM1, the framework captures large-scale discharge regimes over a majority of the validated area, with 62% of validated basins achieving acceptable KGE_{drought} performance. Regional disparities reflect both the heterogeneity of GCM skill and the structural limitations identified through the calibration analysis, particularly the misrepresentation of surface–subsurface runoff partitioning.

By anchoring hydrological projections within the IPCC socioeconomic framework, HyIR provides a consistent basis for
555 prospective water availability assessment across contrasting emission scenarios. In contrast to many climate impact assessments that ultimately report river discharge, HyIR provides explicit storage accounting across rivers, lakes and reservoirs, and groundwater storage compartments, enabling source-specific freshwater availability indicators.

HyIR is not intended to replace global hydrological models. Instead, it provides a lightweight, scenario-native aggregation framework that transforms CMIP6 climate projections into transparent, source-specific freshwater availability indicators
560 suitable for long-term resilience assessment across a wide range of water-dependent sectors.

<https://doi.org/10.5194/egusphere-2026-4606>

Preprint. Discussion started: 6 October 2026

© Author(s) 2026. CC BY 4.0 License.



Appendix A: List of symbols



Table A1. List of symbols, units, and descriptions.

Symbol	Unit	Description
<i>CMIP6 climate forcing</i>		
P	$\text{m}^3 \text{s}^{-1}$	Total precipitation at HydroUnit scale
RET	$\text{m}^3 \text{s}^{-1}$	Real evapotranspiration at HydroUnit scale
R	$\text{m}^3 \text{s}^{-1}$	Total runoff from GCM
R_s	$\text{m}^3 \text{s}^{-1}$	Surface runoff from GCM
F	$\text{kg m}^{-2} \text{s}^{-1}$	Raw CMIP6 flux variable
ρ	kg m^{-3}	Water density (999.5 kg m^{-3})
<i>Water balance</i>		
P_{eff}	$\text{m}^3 \text{s}^{-1}$	Effective precipitation (net of evapotranspiration)
ϕ_s	$\text{m}^3 \text{s}^{-1}$	Surface runoff flux ($= R_s$)
ϕ_b	$\text{m}^3 \text{s}^{-1}$	Baseflow flux ($= R - R_s$)
ϕ_{inf}	$\text{m}^3 \text{s}^{-1}$	Infiltration flux
A_{HU}	m^2	Continental land area of a HydroUnit
ΔV_{SWB}	m^3	Total surface water body storage change in a HydroUnit
ΔV_{aq}	m^3	Groundwater storage variation in a HydroUnit
$Q_{\text{riv},\text{in}}$	$\text{m}^3 \text{s}^{-1}$	Total lateral inflow generated by a HydroUnit
AI	–	Aridity index (ratio of mean annual precipitation to potential evapotranspiration)
$AI_{\text{threshold}}$	–	Aridity classification threshold (default: 0.65)
<i>Surface water body storage</i>		
V_{swb}^i	m^3	Volume of surface water body i
$V_{\text{swb-max}}^i$	m^3	Maximum storage capacity of water body i
θ_{init}	–	Initial volume fraction of maximum capacity (default: 0.40)
A_{swb}^i	m^2	Area of surface water body i
A_{min}	m^2	Area of the smallest water body in the dataset
θ	–	Area intersection threshold for water body inclusion (default: 0.10)
$\bar{d}$	m	Global mean lake depth (3.9 m)
$P_{\text{eff} \rightarrow \text{swb}}^i$	$\text{m}^3 \text{s}^{-1}$	Effective precipitation falling on water body i
$\phi_{s \rightarrow \text{swb}}^i$	$\text{m}^3 \text{s}^{-1}$	Surface runoff entering water body i
$\phi_{\text{swb} \rightarrow \text{aq}}^i$	$\text{m}^3 \text{s}^{-1}$	Exchange flux between water body i and groundwater (positive: downward; negative: upward)



Symbol	Unit	Description (continued)
$\phi_{\text{swb} \rightarrow \text{riv}}^i$	$\text{m}^3 \text{s}^{-1}$	Drainage flux from water body i to the river network
K_{sat}^i	m s^{-1}	Effective exchange conductance beneath water body i
ω_{sat}^i	–	Saturation weighting factor
∇h^i	–	Prescribed large-scale hydraulic exchange gradient (10^{-3})
C^i	s^{-1}	Drainage coefficient of water body i ($\approx 2.96 \times 10^{-7} \text{s}^{-1}$)
$\delta^i(t)$	–	Seasonal modulation factor
$\bar{\delta}$	–	Mean seasonal modulation ($= 1/12$)
A_{δ}	–	Seasonal amplitude ($= 0.02$)
φ	months	Hemisphere-aware phase shift of seasonal drainage
f_{swb}	–	Fractional area coverage of surface water bodies in a HydroUnit
<i>Groundwater storage</i>		
V_{aq}	m^3	Groundwater storage volume
$V_{\text{aq-max}}$	m^3	Maximum groundwater storage capacity
$V_{\text{aq,init}}$	m^3	Initial groundwater storage volume
d_{avg}	m	Area-weighted mean depth to bedrock of a storage unit
A_{aq}	m^2	Area of a groundwater storage unit
n_e	–	Effective porosity
$\phi_{\text{aq} \rightarrow \text{riv}}$	$\text{m}^3 \text{s}^{-1}$	Baseflow contribution from groundwater to the river network
<i>River network routing</i>		
F^i	$\text{m}^3 \text{s}^{-1}$	Lateral inflow to reach i
Q_{in}^i	$\text{m}^3 \text{s}^{-1}$	Upstream inflow to reach i
Q_{out}^i	$\text{m}^3 \text{s}^{-1}$	Outflow from reach i
k^i	s	Muskingum travel time constant for reach i
α^i	–	Muskingum weighting parameter (default: 0.25)
θ	–	Semi-implicit temporal scheme factor (default: 0.5)
A^i, B^i, C^i	–	Muskingum routing coefficients for reach i
$\mathbf{Q}$	$\text{m}^3 \text{s}^{-1}$	Vector of reach outflows
$\mathbf{N}$	–	River network connectivity matrix
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	–	Diagonal matrices of Muskingum coefficients
<i>Parameterization</i>		



Symbol	Unit	Description (continued)
T_c	s	Time of concentration
$T_{c,Pasini}$	s	Time of concentration from Pasini formula
$T_{c,Ventura}$	s	Time of concentration from Ventura formula
F_p	s km^{-1}	Pasini empirical coefficient (3888 s km^{-1})
F_v	s km^{-1}	Ventura empirical coefficient (4578 s km^{-1})
L_{riv}	km	Cumulative upstream river length
A_{bv}	km^2	Upstream watershed area
I	m m^{-1}	Mean watershed slope
$I_{tr}(i)$	–	Relative transfer time index for reach i
$I_{tr,max}$	–	Sum of transfer time indices in a basin
dl_i	m	Length of reach i
$z(i), z_{ds}(i)$	m	Elevation of reach i and its downstream neighbor
$A_{up}(i)$	m^2	Upstream drainage area of reach i
β	–	Transfer time index exponent (0.25)
K_{sat}	m s^{-1}	Saturated hydraulic conductivity
ω_{sat}	–	Soil porosity at saturation
ρ_b	g cm^{-3}	Soil bulk density
ρ_p	g cm^{-3}	Soil particle density (2.65 g cm^{-3})
<i>Calibration parameters</i>		
γ_{T_c}	–	Time-of-concentration corrector $\in (0, 1]$ (default: 1)
γ_{inf}	–	Infiltration-to-runoff redirector $\in (-1, 1)$ (default: 0)
γ_b	–	Baseflow seasonal corrector $\in \mathbb{R}$ (default: 0)
$w(t)$	–	Hemisphere-aware seasonal weight $\in [0, 1]$
<i>Validation metrics</i>		
r	–	Pearson correlation coefficient between simulated and observed discharge
α	–	KGE variability ratio (σ_s/σ_o)
β	–	KGE bias ratio (μ_s/μ_o)
σ_s, σ_o	$\text{m}^3 \text{ s}^{-1}$	Standard deviation of simulated and observed discharge
μ_s, μ_o	$\text{m}^3 \text{ s}^{-1}$	Mean simulated and observed discharge
$QMNA5_s$	$\text{m}^3 \text{ s}^{-1}$	Simulated 5-year return period of annual minimum monthly discharge
$QMNA5_o$	$\text{m}^3 \text{ s}^{-1}$	Observed 5-year return period of annual minimum monthly discharge



565	Symbol	Unit	Description (continued)
-----	--------	------	-------------------------

<https://doi.org/10.5194/egusphere-2026-4606>

Preprint. Discussion started: 6 October 2026

© Author(s) 2026. CC BY 4.0 License.



Appendix B: List of abbreviations



Table B1. List of symbols, units, and descriptions.

Abbreviation	Definition
<i>Climate and scenario frameworks</i>	
CMIP6	Coupled Model Intercomparison Project Phase 6
GCM	General Circulation Model
ESM	Earth System Model
ESGF	Earth System Grid Federation
IPCC	Intergovernmental Panel on Climate Change
IPCC AR6	IPCC Sixth Assessment Report
SSP	Shared Socioeconomic Pathway
ERA5	Fifth generation ECMWF atmospheric reanalysis
<i>Hydrological framework</i>	
HyIR	HYdrological and Industrial Resilience
GHM	Global Hydrological Model
ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
<i>Spatial and computational</i>	
EPSG	European Petroleum Survey Group (coordinate reference system registry)
SRTM	Shuttle Radar Topography Mission
<i>Input datasets</i>	
GLWD	Global Lakes and Wetlands Database
GLHYMPS	Global Hydrogeology and Permeability of Soils and Sedimentary Basins
GLiM	Global Lithological Map
DTB	Depth to Bedrock
WHYMAP	World Hydrogeological Map
AI	Aridity Index
UNEP	United Nations Environment Programme
<i>River network datasets</i>	
HydroSHEDS	Hydrological data and maps based on SHuttle Elevation Derivatives
HydroRIVERS	Global river network derived from HydroSHEDS
HydroBASINS	Global watershed boundaries derived from HydroSHEDS



Abbreviation	Definition (continued)
RiverATLAS	Global river reach attributes dataset from HydroSHEDS
<i>Validation datasets</i>	
GRDC	Global Runoff Data Centre
GloFAS	Global Flood Awareness System
<i>Performance metrics</i>	
KGE	Kling–Gupta Efficiency
KGE _{drought}	Drought-focused Kling–Gupta Efficiency
QMNA5	5-year return period of the annual minimum monthly mean discharge
<i>Soil and hydrogeology</i>	
SoilGrids	Global gridded soil information database (ISRIC)
ISRIC	International Soil Reference and Information Centre



Appendix C: Input datasets

Dataset	Content	Source institution	Access
<i>Climate forcing</i>			
CMIP6: TaiESM1 outputs	Monthly precipitation, evapotranspiration, surface runoff, and total runoff under historical and SSP scenarios	World Climate Research Programme (CMIP6); Research Center for Environmental Changes, Taiwan (TaiESM1)	https://esgf.llnl.gov
<i>Climate zone classification</i>			
Future Global Aridity Index and PET Database	Historical and GCM- and SSP-specific future aridity index fields	Zomer and Trabucco (2025)	https://doi.org/10.57760/sciencedb.nbsdc.00086
<i>Spatial datasets</i>			
GLWD	Lake and reservoir polygons with area and volume attributes	WWF and Center for Environmental Systems Research, University of Kassel	https://www.worldwildlife.org/our-work/science/global-lakes-and-wetlands-database
Depth to Bedrock	Global spatially continuous depth-to-bedrock estimates at 10 km resolution	Shangguan et al. (2017)	https://soilgrids.org
GLHYMPS	Global hydrogeology and permeability of soils and sedimentary basins	Gleeson (2018)	https://doi.org/10.1002/2014GL059856
SoilGrids	Global gridded soil bulk density at multiple depth intervals	ISRIC – World Soil Information	https://soilgrids.org
GLiM	Global Lithological Map for gap-filling of hydraulic conductivity	Hartmann and Moosdorf (2012)	https://doi.org/10.1029/2012GC004370
Surya et al. (2020)	Global measured and modeled saturated hydraulic conductivity	Surya et al. (2020)	https://doi.org/10.1002/vzj2.20012



Dataset	Content	Source institution	Access
HydroSHEDS: HydroRIVERS, HydroBASINS	Global river network, watershed boundaries at Pfafstetter levels 1–12, and reach at-tributes	HydroSHEDS	https://www.hydrosheds.org
<i>Observed discharge for validation</i>			
GRDC	Daily river discharge records from global gauge stations	Global Runoff Data Centre (GRDC)	https://grdc.bafg.de
GloFAS reanalysis	Daily gridded discharge at 0.1° resolution, ERA5-forced, 1979–2014	ECMWF / Copernicus Climate Change Service	https://cds.climate.copernicus.eu

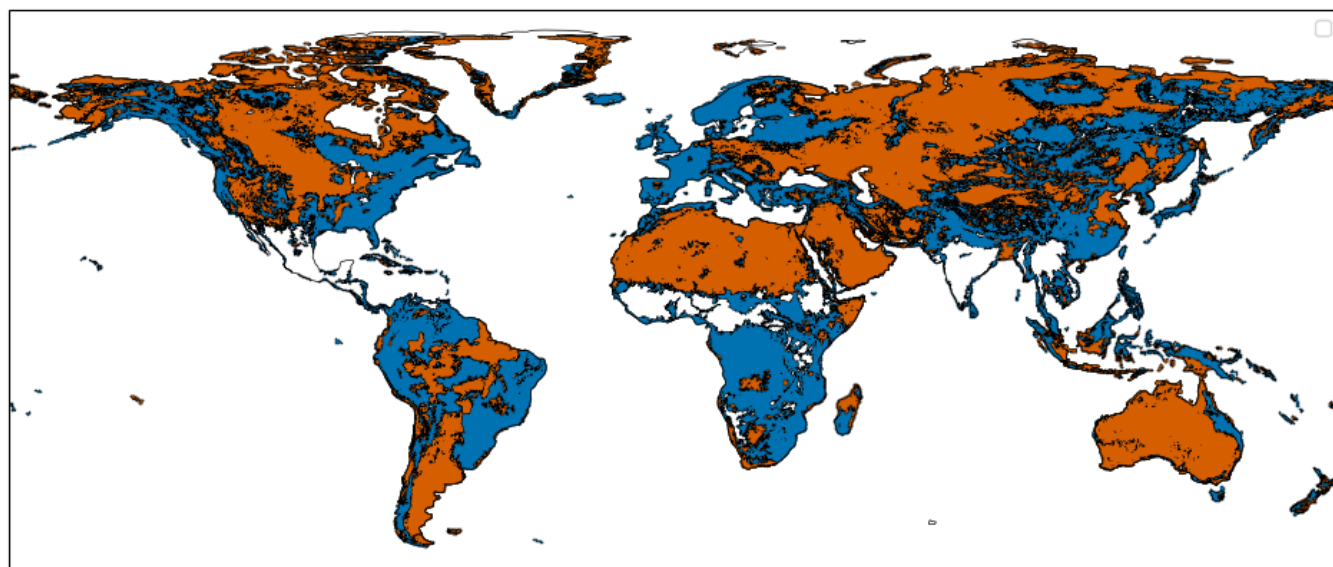


Figure D1. Global delineation of groundwater storage units derived from the depth-to-bedrock dataset of Shangguan et al. (2017). Orange: deep storage units (depth > 20 m); blue: shallow storage units (depth 5–20 m). White areas correspond to regions where depth to bedrock is less than 5 m, which are excluded from the groundwater storage compartment, or to oceanic and Antarctic cells outside the HyIR domain.

570 Appendix D: Global groundwater storage unit delineation

Appendix E: Key dataset descriptions

E1 CMIP6

The Coupled Model Intercomparison Project Phase 6 (CMIP6) is a coordinated framework of climate model experiments organized under the World Climate Research Programme (WCRP). It brings together output from over 100 global climate models developed by approximately 50 modeling institutions worldwide, run under a common set of forcing protocols covering the historical period (1850–2014) and future projections (2015–2100) under standardized Shared Socioeconomic Pathway (SSP) scenarios. CMIP6 constitutes the primary observational and projection basis for the IPCC Sixth Assessment Report (Eyring et al., 2016; Masson-Delmotte et al., 2021).

E2 TaiESM1

580 TaiESM1 (Taiwan Earth System Model version 1) is a global coupled climate model developed by the Research Center for Environmental Changes, Academia Sinica, Taiwan (AS-RCEC), based on the Community Earth System Model version 1.2.2 (Lee et al., 2020). It participated in CMIP6 and provides historical simulations (1850–2014) and future projections under multiple SSP scenarios at a native atmospheric resolution of $1.25^\circ \times 0.9^\circ$.



E3 GRDC

585 The Global Runoff Data Centre (GRDC) is an international data centre operating under the auspices of the World Meteorological Organization (WMO), hosted by the Federal Institute of Hydrology (BfG) in Koblenz, Germany. It maintains the most comprehensive global archive of quality-assured daily and monthly river discharge records from gauge stations worldwide, covering records from the 19th century to the present (The Global Runoff Data Centre (GRDC), 2025).

E4 GloFAS

590 The Global Flood Awareness System (GloFAS) reanalysis is a gridded daily river discharge dataset at 0.1° resolution, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Climate Change Service. It is derived from the Lisflood hydrological model forced with ERA5 atmospheric reanalysis, covering the period 1979 to present, and provides spatially complete discharge estimates for large river systems globally (Harrigan et al., 2020).



Code and data availability. The current version of HyIR 0.1 is available from the project repository at <https://gitlab.com/hyir/Hyir> under the
595 Eclipse Public License v2.0. The exact version of the code used to produce the results presented in this paper is archived on Zenodo at <https://doi.org/10.5281/zenodo.21701474> (Taki et al., 2026). The platform is implemented in Python following an object-oriented architecture. All
framework parameters are externally configurable through a structured configuration file.

The third-party datasets used to force and evaluate the model are subject to third-party licensing and redistribution restrictions and could
therefore not be archived alongside the code. They remain publicly accessible from their original providers: CMIP6 climate forcing data
600 are available through the Earth System Grid Federation (ESGF) at <https://esgf.llnl.gov>; HydroRIVERS and HydroBASINS are available
from HydroSHEDS (<https://www.hydrosheds.org>); GLWD data are available from <https://www.worldwildlife.org>; GRDC discharge data
are available from <https://grdc.bafg.de>; and GloFAS reanalysis data are available from the Copernicus Climate Change Service at <https://cds.climate.copernicus.eu>. Further details on dataset versions and access are provided in Appendix C.

Author contributions. M.T. designed and implemented HyIR, performed the simulations, calibration, validation and visualization, and led
605 the writing of the manuscript. N.F. led the conceptual design of the hydrological framework, provided scientific and software supervision
throughout the development, and contributed to the scientific interpretation of results. D.G. contributed to the industrial application design
and provided scientific supervision. All authors reviewed and edited the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Financial support. This work was carried out within the department of Geosciences and Geoengineering at Mines Paris – PSL University,
610 under the research chair “Mineral Resources for the Energy Transition”, signed by EMINES – Mohamed VI Polytechnical University (UM6P)
and Mines Paris – PSL University, with the support of the OCP Group.

Acknowledgements. The authors thank the CMIP6 modeling groups, the Program for Climate Model Diagnosis and Intercomparison and the
WCRP Working Group on Coupled Modelling for their roles in making the WCRP CMIP6 multi-model dataset available. The authors also
thank the HydroSHEDS development team for providing the HydroRIVERS and HydroBASINS datasets; the World Wildlife Fund for the
615 Global Lakes and Wetlands Database (GLWD); the Global Runoff Data Centre (GRDC, Koblenz, Germany) for providing river discharge
data; and the Copernicus Climate Change Service for providing the GloFAS reanalysis data.



References

- Alcamo, J., Flörke, M., and Märker, M.: Future long-term changes in global water resources driven by socio-economic and climatic changes, *Hydrological Sciences Journal*, 52, 247–275, <https://doi.org/10.1623/hysj.52.2.247>, 2007.
- 620 Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, *Water Resources Research*, 51, 4923–4947, <https://doi.org/10.1002/2015WR017173>, 2015.
- Cunge, J. A.: On the subject of a flood propagation computation method (Muskingum method), *Journal of Hydraulic Research*, 7, 205–230, <https://doi.org/10.1080/00221686909500264>, 1969.
- David, C. H., Yang, Z.-L., and Hong, S.: Regional-scale river flow modeling using off-the-shelf runoff products, thousands of mapped rivers
625 and hundreds of stream flow gauges, *Environmental Modelling & Software*, 42, 116–132, <https://doi.org/10.1016/j.envsoft.2012.12.011>, 2013.
- de Almeida, I. K., Almeida, A. K., Anache, J. A. A., Steffen, J. L., and Alves Sobrinho, T.: Estimation on time of concentration of overland flow in watersheds: A review, *Geociências*, 33, 661–671, 2014.
- De Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., and Bierkens, M. F. P.: A high-resolution global-scale groundwater model,
630 *Hydrology and Earth System Sciences*, 19, 823–837, <https://doi.org/10.5194/hess-19-823-2015>, 2015.
- Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, 270, 105–134, [https://doi.org/10.1016/S0022-1694\(02\)00283-4](https://doi.org/10.1016/S0022-1694(02)00283-4), 2003.
- Döll, P., Trautmann, T., Gerten, D., Müller Schmied, H., Ostberg, S., Saaed, F., and Schleussner, C.-F.: Risks for the global freshwater system at 1.5 °C and 2 °C global warming, *Environmental Research Letters*, 13, 044 038, <https://doi.org/10.1088/1748-9326/aab792>, 2018.
- 635 ESGF: ESGF MetaGrid, <https://esgf.llnl.gov>, 2026.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Flipo, N., Monteil, C., Poulin, M., de Fouquet, C., and Krimissa, M.: Hybrid fitting of a hydrosystem model: Long-term insight into the
640 Beauce aquifer functioning (France), *Water Resources Research*, 48, <https://doi.org/10.1029/2011WR011092>, 2012.
- Flipo, N., Gallois, N., and Schuite, J.: Regional coupled surface–subsurface hydrological model fitting based on a spatially distributed minimalist reduction of frequency domain discharge data, *Geoscientific Model Development*, 16, 353–381, <https://doi.org/10.5194/gmd-16-353-2023>, 2023.
- Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., Geiger, T.,
645 Halladay, K., Hurtt, G., Mengel, M., Murakami, D., Ostberg, S., Popp, A., Riva, R., Stevanovic, M., Suzuki, T., Volkholz, J., Burke, E., Ciais, P., Ebi, K., Eddy, T. D., Elliott, J., Galbraith, E., Gosling, S. N., Hattermann, F., Hickler, T., Hinkel, J., Hof, C., Huber, V., Jägermeyr, J., Krysanova, V., Marcé, R., Müller Schmied, H., Mouratiadou, I., Pierson, D., Tittensor, D. P., Vautard, R., van Vliet, M., Biber, M. F., Betts, R. A., Bodirsky, B. L., Deryng, D., Froliking, S., Jones, C. D., Lotze, H. K., Lotze-Campen, H., Sahajpal, R., Thonicke, K., Tian, H., and Yamagata, Y.: Assessing the impacts of 1.5 °C global warming above pre-industrial levels – an ISIMIP2b approach, *Geoscientific Model Development*, 10, 4321–4345, <https://doi.org/10.5194/gmd-10-4321-2017>, 2017.
- 650 Gleeson, T.: GLObal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, <https://doi.org/10.20383/101.0111>, 2018.
- Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T. I. E., Vetter, T., Wang, X., and Zhang, X.: A comparison of changes in river runoff



- from multiple global and catchment hydrological models under global warming scenarios of 1 °C, 2 °C and 3 °C, *Climatic Change*, 141, 577–595, <https://doi.org/10.1007/s10584-016-1773-3>, 2017.
- Grimaldi, S., Petroselli, A., Tauro, F., and Porfiri, M.: Time of concentration: a paradox in modern hydrology, *Hydrological Sciences Journal*, 57, 217–228, <https://doi.org/10.1080/02626667.2011.644244>, 2012.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- 660 Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z. D., Wada, Y., and Wisser, D.: Global water resources affected by human interventions and climate change, *Proceedings of the National Academy of Sciences*, 111, 3251–3256, <https://doi.org/10.1073/pnas.1222475110>, 2014.
- Hamed, F., Bozorg-Haddad, O., Pazoki, M., Asgari, H.-R., Parsa, M., and Loáiciga, H. A.: Parameter estimation of extended non-linear Muskingum models with the weed optimization algorithm, *Journal of Irrigation and Drainage Engineering*, 142, 04016059, [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0001095](https://doi.org/10.1061/(ASCE)IR.1943-4774.0001095), 2016.
- 665 Harrigan, S., Zsoter, E., Cloke, H., Salamon, P., and Prudhomme, C.: Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System, *Hydrology and Earth System Sciences*, 24, 2821–2840, <https://doi.org/10.5194/hess-24-2821-2020>, 2020.
- Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, *Geochemistry, Geophysics, Geosystems*, 13, <https://doi.org/10.1029/2012GC004370>, 2012.
- 670 Hillel, D.: *Environmental Soil Physics*, Academic Press, 1st edn., 1998.
- Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, *Nature Climate Change*, 3, 816–821, <https://doi.org/10.1038/nclimate1911>, 2013.
- Hou, Y., Guo, H., Yang, Y., and Liu, W.: Global evaluation of runoff simulation from climate, hydrological and land surface models, *Water Resources Research*, 59, e2021WR031817, <https://doi.org/10.1029/2021WR031817>, 2023.
- 675 Huntington, T. G.: Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, 319, 83–95, <https://doi.org/10.1016/j.jhydrol.2005.07.003>, 2006.
- International Council on Mining and Metals (ICMM): *Water management in mining: A selection of case studies*, <https://www.icmm.com>, 2012.
- 680 ISRIC – World Soil Information: *SoilGrids – global gridded soil information*, <https://soilgrids.org>, 2020.
- 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.
- Korkmaz, S.: *Modeling of the flood regimes in coupled stream-aquifer systems*, PhD thesis, 2007.
- Lee, H. and Romero, J., eds.: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva, Switzerland, <https://doi.org/10.59327/IPCC/AR6-9789291691647>, 2023.
- 685 Lee, W.-L., Wang, Y.-C., Shiu, C.-J., Tsai, I.-C., Tu, C.-Y., Lan, Y.-Y., Chen, J.-P., Pan, H.-L., and Hsu, H.-H.: Taiwan Earth System Model Version 1: Description and evaluation of mean state, *Geoscientific Model Development*, 13, 3887–3904, <https://doi.org/10.5194/gmd-13-3887-2020>, 2020.
- 690 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.
- 695 Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres*, 99, 14 415–14 428, <https://doi.org/10.1029/94JD00483>, 1994.
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., eds.: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, <https://doi.org/10.1017/9781009157896>, 2021.
- 700 Messenger, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, *Nature Communications*, 7, 13 603, <https://doi.org/10.1038/ncomms13603>, 2016.
- 705 Mudd, G. M.: The environmental sustainability of mining in Australia: key mega-trends and looming constraints, *Resources Policy*, 35, 98–115, <https://doi.org/10.1016/j.resourpol.2009.12.001>, 2010.
- Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, *Geoscientific Model Development*, 14, 1037–1079, <https://doi.org/10.5194/gmd-14-1037-2021>, 2021.
- 710 Northey, S. A., Mudd, G. M., and Werner, T. T.: Unresolved complexity in assessments of mineral resource depletion and availability, *Natural Resources Research*, 27, 241–255, <https://doi.org/10.1007/s11053-017-9352-5>, 2017.
- Pasini, F.: *Relazione sul progetto della bonifica renana*, 1914.
- Richts, A., Struckmeier, W. F., and Zaepke, M.: WHYMAP and the groundwater resources map of the world 1:25,000,000, in: *Sustaining Groundwater Resources*, edited by Jones, J. A. A., pp. 159–173, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-90-481-3426-7_10, 2011.
- 715 Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. A.: A safe operating space for humanity, *Nature*, 461, 472–475, <https://doi.org/10.1038/461472a>, 2009.
- 720 Sacks, W. J., Cook, B. I., Buening, N., Levis, S., and Helkowski, J. H.: Effects of global irrigation on the near-surface climate, *Climate Dynamics*, 33, 159–175, <https://doi.org/10.1007/s00382-008-0445-z>, 2010.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, 46, <https://doi.org/10.1029/2008WR007327>, 2010.
- Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, *Proceedings of the National Academy of Sciences*, 111, 3245–3250, <https://doi.org/10.1073/pnas.1222460110>, 2014.
- 725 Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in:



- 730 Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC, edited by Masson-Delmotte, V. et al., pp. 1513–1766, Cambridge University Press, <https://doi.org/10.1017/9781009157896.013>, 2021.
- Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, *Journal of Advances in Modeling Earth Systems*, 9, 65–88, <https://doi.org/10.1002/2016MS000686>, 2017.
- Surya, G., Hengl, T., Lehmann, P., Bonetti, S., and Or, D.: SoilKsatDB: global compilation of soil saturated hydraulic conductivity measure-
735 ments for geoscience applications, *Earth System Science Data*, 12, 1593–1612, <https://doi.org/10.5194/essd-12-1593-2020>, 2020.
- 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., Karssenber, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannamettee, E., Wissner, 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.
- 740 Taki, M., Flipo, N., and Goetz, D.: HyIR 0.1: A Global Hydrological Aggregation Framework for Translating CMIP6 Climate Projections into Freshwater Availability Scenarios, <https://doi.org/10.5281/zenodo.21701474>, 2026.
- Tallaksen, L. M. and van Lanen, H. A. J., eds.: *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*, vol. 48 of *Developments in Water Science*, Elsevier, Amsterdam, 2004.
- Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M.,
745 Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground water and climate change, *Nature Climate Change*, 3, 322–329, <https://doi.org/10.1038/nclimate1744>, 2013.
- The Global Runoff Data Centre (GRDC): Global Runoff Data Centre – WaterData, <https://grdc.bafg.de>, 2025.
- United Nations Environment Programme (UNEP): *World Atlas of Desertification*, 1992.
- 750 United Nations World Water Assessment Programme (WWAP): *The United Nations World Water Development Report 2019: Leaving No One Behind*, 2019.
- Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Müller Schmied, H., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, *Environmental Research Letters*, 13, 055 008, <https://doi.org/10.1088/1748-9326/aab96f>, 2018.
- 755 Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, *Nature*, 467, 555–561, <https://doi.org/10.1038/nature09440>, 2010.
- Wada, Y. and Bierkens, M. F. P.: Sustainability of global water use: past reconstruction and future projections, *Environmental Research Letters*, 9, 104 003, <https://doi.org/10.1088/1748-9326/9/10/104003>, 2014.
- 760 Zomer, R. and Trabucco, A.: Future Global Aridity Index and PET Database (CMIP6), <https://doi.org/10.6084/m9.figshare.7504448>, 2025.