# How well do hydrological models simulate streamflow extremes and drought-to-flood transitions?

Eduardo Muñoz-Castro[1,2,3], Bailey J. Anderson[1,2,3], Paul C. Astagneau[1,2,3], Daniel L. Swain[4,5,6], Pablo A. Mendoza[7,8], and Manuela I. Brunner[1,2,3]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos Dorf, Switzerland
[2]Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos Dorf, Switzerland
[3]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[4]California Institute for Water Resources, University of California Agriculture and Natural Resources, Davis, CA, USA
[5]Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, USA
[6]Capacity Center for Climate and Weather Extremes, National Center for Atmospheric Research, Boulder, CO, USA
[7]Civil Engineering Department, Universidad de Chile, Santiago, Chile
[8]Advanced Mining Technology Centre (AMTC), Universidad de Chile, Santiago, Chile

**Correspondence:** Eduardo Muñoz-Castro (eduardo.munoz-castro@slf.ch)

**Abstract.** Flood impacts can be enhanced when they occur shortly after droughts. Hydrological models are useful tools to better understand the underlying processes and mechanisms driving the response of floods occurring in close succession to streamflow drought. However, it is yet unclear how well hydrological models capture these compound extreme events and which modeling decisions are most important for high model performance. To address this research gap, we calibrated four conceptual bucket-style hydrological models with different structures (GR4J, GR5J, GR6J, and TUW) for 63 catchments located in Chile and Switzerland using different calibration strategies. Specifically, we assessed the relative importance of different methodological choices in simulating and detecting observed drought-to-flood transitions including model structure, streamflow transformation, and the Kling-Gupta efficiency (KGE) formulation and weights used to calibrate the model parameters. We demonstrate that model performance as expressed by the KGE does not guarantee a good performance in terms of detecting streamflow extremes and their transitions. Further, we show that a model's performance with respect to capturing extreme events primarily depends on how well it captures streamflow timing. Our results also highlight that model structure, catchment characteristics and meteorological forcings play a key role in the detection of transitions. Specifically, we demonstrate that drought-to-flood transitions are more difficult to capture in semi-arid high-mountain catchments than in humid low-elevation catchments. Ultimately, our study provides insights for further model improvements to simulate and better understand drought-to-flood transitions and to identify regions prone to this hazard.

## 1 Introduction

Hydrological extreme events such as streamflow droughts and floods are expected to become more frequent, severe, and persistent in a warming climate (e.g., Gu et al., 2023; Asadieh and Krakauer, 2017; Martin, 2018; Tabari et al., 2021). Therefore, severe impacts are expected on infrastructure, agriculture, water supply, and hydropower generation (e.g., McClymont et al., 2020; McMartin et al., 2018; Lehner et al., 2006; Sivakumar, 2011; Wasti et al., 2022), as well as social and political systems (e.g., Doocy et al., 2013; Hurlbert and Gupta, 2017; Kiem and Austin, 2013; Visconti, 2022).

Studies focusing on hydrological extreme events and their impacts often assume temporal and/or spatial independence between them, neglecting that extremes may be multivariate phenomena (Banfi and De Michele, 2022; Brunner, 2023). However, the impacts of floods can be enhanced when they occur during or shortly after a streamflow drought (e.g., Barendrecht et al., 2024; Swain et al., 2018; He and Sheffield, 2020; Rashid and Wahl, 2022). For instance, Handwerger et al. (2019) and Valenzuela et al. (2022) have demonstrated an increase in the occurrence of landslides in California and Chile due to shifts from meteorological drought to intense precipitation. Similarly, Dietze et al. (2022) have shown that the 2018-2020 drought in Europe has enhanced the debris mobilisation during the 2021 flood in the Eifel region of western Germany and Belgium. In 2017, intense precipitation broke the 2012-2016 drought in California and led to severe flooding, the activation of the emergency spillway of the Lake Oroville dam for the first time in its history, and the declaration of emergency (Griffin and Anchukaitis, 2014; Robeson, 2015; Wang et al., 2017). Despite the need to integrate both streamflow extreme events within the same analysis framework (e.g., Ward et al., 2020; Quesada-Montano et al., 2018; Di Baldassarre et al., 2017), droughts and floods have been mostly studied as independent events.

As a consequence of a potential intensification of hydrological volatility in a warming climate (Swain et al., 2025), hydrological whiplash, defined as sub-seasonal transitions between hydrological extremes such as droughts and floods (Hammond et al., 2025), could become more frequent and severe in the future. While the transition from drought to flood can occur within hours or days, the transition to droughts can range from weeks to years, leading to different water management challenges and reaction times for decision-makers (Hammond et al., 2025). Then, due to the inherent asymmetry in spatiotemporal characteristics and underlying drivers, as has been recently shown by Swain et al. (2025) from both meteorological and hydrological perspectives, drought-to-flood transitions can have more severe impacts than flood-to-drought transitions. Both hydrological droughts and floods are linked to meteorological conditions such as precipitation surplus/deficit or low/high evapotranspiration rates. However, Brunner et al. (2025) have shown that dry-to-wet spells are only weakly associated with drought-to-flood transitions, with a propagation rate of just 10% within a 30-day period, and that wet spells are less likely to lead to floods than dry spells are to cause droughts. Consequently, the occurrence and drivers of these compound events are not yet fully understood (e.g., Matanó et al., 2022, 2024; Brunner, 2023; Götte and Brunner, 2024; Hammond et al., 2025; Brunner et al., 2025).

Process-based hydrological models can provide valuable insights on how streamflow and/or other hydrological fluxes and states react to variations in meteorological and environmental inputs (Hrachowitz and Clark, 2017). In recent decades, several efforts have been made to improve the realism of hydrological models in terms of spatial variability (e.g., Dembélé et al., 2020), the simulation of low (e.g., Garcia et al., 2017) and high flows (e.g., Mizukami et al., 2019), or the representation of flood-triggering mechanisms and spatiotemporal coherence (e.g., Brunner et al., 2020, 2021), under current and changing climatic conditions (e.g., Fowler et al., 2018). However, modeling hydrological extreme events such as droughts and floods is still challenging (e.g., Mizukami et al., 2019; Bruno et al., 2024), especially when multiple variables are involved. Such cases include, for example, modeling the dependence between flood peaks and volumes (Brunner and Sikorska-Senoner, 2019), or modeling the spatial dependence of floods happening in different locations (Brunner et al., 2021). This complexity indicates that capturing consecutive drought-to-flood events might not be trivial either. As model evaluations targeted at compound extremes have not yet been performed, it is still unclear how well hydrological models can, in fact, capture drought-to-flood transitions.

Hydrological modeling involves making decisions about model structure (i.e., process representations and parameterizations), spatial discretization, meteorological forcings, and parameter estimation approach (e.g., calibration/evaluation periods, hydrological target variables or signatures used in objective function), which affect hydrological simulations and whose importance might vary depending on the modeling purpose (e.g., Mendoza et al., 2016; Mizukami et al., 2016; Baez-Villanueva et al., 2021; Guo et al., 2017; Melsen et al., 2019). Previous studies have highlighted that such modeling decisions can substantially influence simulated hydrological extremes and their uncertainties (e.g., Alexander et al., 2023; Melsen and Guse, 2019; Melsen et al., 2019). They have also shown that the choice of objective function for model calibration, model structure, and spatial discretization (forcings and domain) are the most influential decisions on modeling outcomes. While these previous studies have focused on analyzing the impacts of modeling decisions on drought and flood attributes (e.g., severity, duration), they have not looked at how these decisions influence event detection, i.e. whether or not a model can capture extreme events below or above a certain threshold. Moreover, previous work has focused on individual extremes instead of looking at them in a multivariate setting. As such, it is yet unclear how individual modeling decisions might influence the representation of hydrological transitions.

Hydrological modeling often relies on the calibration process to find parameter values that minimize discrepancies between observations and simulations of a target variable (e.g., streamflow). The calibration process requires defining an objective function to measure the similarity between observations and simulations. In general, these objective functions are defined based on "least squares" formulations such as the widely used Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) and the Kling-Gupta Efficiency (KGE; Gupta et al., 2009). Although alternative objective functions have been proposed to enhance the robustness of calibrated parameters and hydrological consistency (e.g., Fowler et al., 2018; Yilmaz et al., 2008; McMillan, 2020), KGE and NSE remain widely used for model calibration and evaluation (e.g., Klemeš, 1986; Motavita et al., 2019; Seibert et al., 2019; Beven, 2025).

The Kling-Gupta Efficiency (KGE), originally proposed by Gupta et al. (2009), has been one of the most popular performance metrics used in hydrology over the last decades. Thanks to the possibility of disaggregating it into its three components — bias, variability, and correlation — KGE provides interpretability and diagnostic power. It has been applied for many modeling purposes, including the analysis of streamflow extremes (e.g., Gu et al., 2023; Hirpa et al., 2018). In these studies, despite the lack of objectivity in assessing the model's explanatory power in each catchment (i.e., benchmark; e.g., Knoben, 2024; Seibert et al., 2018), calibrations are considered successful if the KGE performance exceeds a certain threshold during both the calibration and evaluation periods (e.g., KGE > 0.4). This criterion is used as a proxy for how well a model represents streamflow properties such as extreme events (e.g., Lema et al., 2025; Cinkus et al., 2023; Zhao et al., 2025). However, there is often no explicit evaluation of how drought or flood events are represented at the event scale. Thus, the suitability of KGE and alternative formulations (Gupta et al., 2009; Kling et al., 2012; Pool et al., 2018; Tang et al., 2021; Pizarro and Jorquera, 2024) or adaptations (e.g., transformations and weights; Garcia et al., 2017; Wu et al., 2025; Mizukami et al., 2019) for calibrating models aimed at studying streamflow extreme events and, in particular, consecutive extremes, has not yet been sufficiently evaluated.

In summary, the effectiveness of overall performance metrics - such as KGE - in evaluating the models' ability to capture streamflow extremes has not yet been thoroughly examined. Additionally, it is unclear how different modeling decisions - such as the hydrological model, objective function, and streamflow transformations - affect drought-to-flood transition simulations. Even more, it remains to be explored which modeling choices are most suitable for capturing these compound hydrological extreme events without compromising hydrological consistency (i.e., representation of different hydrological processes or properties). To address these research gaps, we investigate the extent to which hydrological models can represent consecutive drought-to-flood transitions and the impact of model structure and calibration choices on their representation. Specifically, we address the following research questions:

- How suitable is the KGE for calibrating models aimed at jointly simulating streamflow droughts and floods?
- Which modeling choices (e.g., model structure, KGE formulation, etc.) are most important for simulating droughts, floods, and their transitions?
- Which are the key hydrological processes that have to be captured by models to simulate drought-to-flood transitions well?

To address these questions, we performed several calibration experiments with four conceptual bucket-type hydrological models (GR4J, GR5J, GR6J, and TUW) across 63 catchments in Chile and Switzerland. In our experiments, we tested different configurations of the Kling-Gupta efficiency (KGE) to assess their performance in simulating and detecting observed transitions. These configurations included five KGE formulations (Table 1), two streamflow transformations (i.e., Q and 1/Q) and their linear combination (i.e., 0.5*KGE(Q)+0.5*KGE(1/Q)), and four different weights applied to the variability term of the KGE ($c_2$=1,2,4,8). Secondly, we assessed the relative importance of each methodological choice for detecting events and ensuring hydrological consistency. Finally, we explored the link between model performance and the representation of different hydrological fluxes and states during transition events.

## 2 Study domain and data

The study domain encompasses 24 and 39 near-natural catchments in Chile (CL; Figure 1a) and Switzerland (CH; Figure 1b), respectively. These catchments are selected based on the availability of complete daily streamflow records between 1981 and 2020 for at least 30 years, with a complete year being defined as one in which all months had information for at least 90% of the days. The selected catchments span a wide range of hydroclimatic characteristics (Figure 1c), from energy to water-limited, and different hydrological regimes (Figure 1d), from snowmelt (e.g., p-seasonality < -0.5 and q-seasonality > 0.5) to rainfall-dominated (e.g., p-seasonality < -0.5 and q-seasonality < -0.5). Some catchments are positioned above the water limit (i.e., Q/P = 1) or below the energy limit (i.e., Q/P = 1 - 1/(P/PET); Figure 1c), which suggests an underestimation of precipitation –

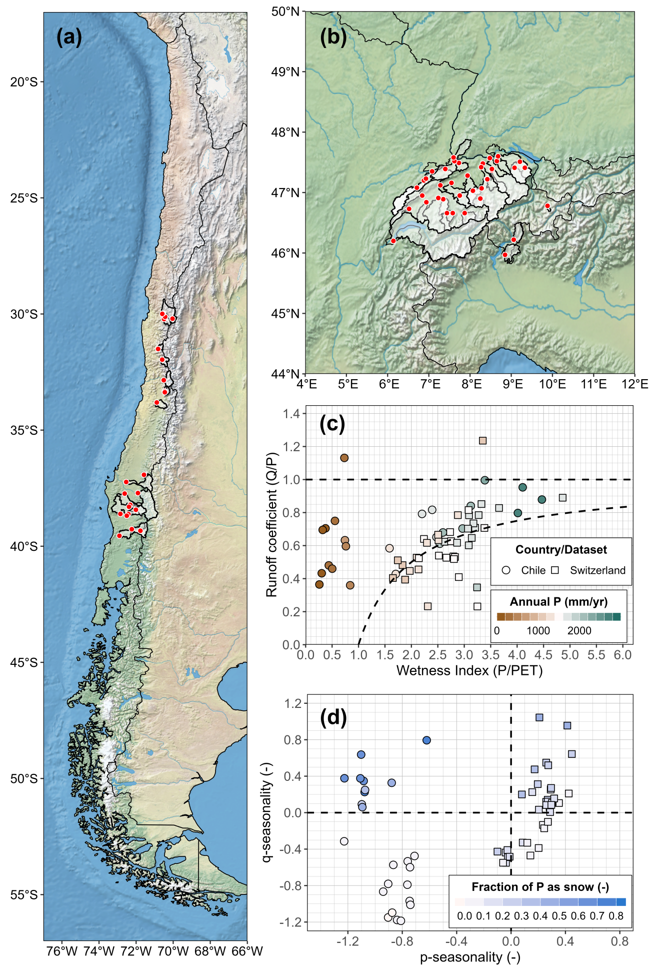


**Figure 1.** Study domain and hydroclimatic characteristics computed for the period 1981-2020 using data retrieved from CAMELS Chile (CL) and Switzerland (CH). Location of catchments across the study domain in (a) Chile and (b) Switzerland, (c) relationship between wetness index (P/PET), runoff coefficient (Q/P), and mean annual precipitation, and (d) relationship of seasonality of precipitation and streamflow and fraction of precipitation falling as snow. For p-seasonality and q-seasonality, positive (negative) values indicate summer (winter) dominated precipitation or streamflow, while values close to zero indicate a uniform distribution across the year.

which might require correcting for precipitation undercatch (e.g., Newman et al., 2015; Stisen et al., 2012; Hughes et al., 2021) – or a surplus of streamflow.

The CAMELS Chile (CL; Alvarez-Garreton et al., 2018a) and Switzerland (CH; Höge et al., 2023a) datasets are used to obtain the meteorological forcings, streamflow records, snow water equivalent (SWE) estimates, and catchment boundaries for the study domains. The meteorological forcings of both datasets, CR2Met version 2.5 for Chile (Boisier, 2023) and RhiresD version 2 for Switzerland (MeteoSwiss, 2023), are based on local gridded observation-based products, while SWE products are based on a snow cover model and data assimilation (for more detail refer to Cortés and Margulis, 2017; Magnusson et al., 2014). We prefer these local products over global ones such as ERA5 (Hersbach et al., 2020) because of their reliance on observations and high horizontal resolutions (approximately 5x5 km$^2$ for CR2Met and 2x2 km$^2$ for RhiresD) that enable a better representation of precipitation patterns in the complex topography of our study domain. Further, these products have been widely used for hydrological studies in Chile (e.g., Vásquez et al., 2021; Alvarez-Garreton et al., 2021; Araya et al., 2023) and Switzerland (e.g., Peleg et al., 2020; Fatichi et al., 2015; Tuel et al., 2022). Streamflow records available through the CAMELS datasets were acquired from the national agencies in each country (i.e., the General Directorate of Water of Chile - DGA and the Swiss Federal Office for the Environment - FOEN). Additionally, we retrieve time series of actual evapotranspiration (ET) and soil moisture (SM) from the satellite and reanalysis-based GLEAM v3.8a dataset (Miralles et al., 2011), spatially aggregated to the catchment scale. We compute topographic characteristics and hypsometric curves - needed to set up the snow routines - using the catchment outlines from CAMELS and the Multi-Error-Removed Improved-Terrain (MERIT) digital elevation model (Yamazaki et al., 2019).

## 3 Methodological approach

Our methodological approach is illustrated in Figure 2. Four hydrological models are calibrated with data from the CAMELS datasets and streamflow records, using five different formulations of the Kling-Gupta efficiency (KGE) as objective functions. In addition, we test three streamflow transformations and four different weights applied to the KGE variability term. This calibration experiment results in 60 optimal parameter sets per model and catchment (i.e., 5 KGE x 3 transformations x 4 weights). We evaluate model performance based on (1) general goodness-of-fit metrics (Legates and McCabe Jr., 1999; Althoff and Rodrigues, 2021), (2) simulation of extreme events and transitions between them, and (3) hydrological consistency in different processes related to streamflow, snow, evapotranspiration, and soil moisture. We also assess model performance at the event scale for droughts, floods, and transitions using categorical indices. In this paper, we use the terms 'formulation' to refer to a specific definition of the KGE (1), 'case' to refer to the application of KGE weights or flow transformations, and 'configuration' to mention the combination of a KGE formulation and a case using certain weights and a specific streamflow transformation. Further, the cases without weights and/or the linear combination between streamflow without (i.e., Q) and with low-flow transformations (i.e., 1/Q) will be used as a reference for the comparison of the results. To assess the statistical significance of the differences between, e.g., the configurations tested in this study, we applied the Wilcoxon test (Wilcoxon, 1945) at a 5% significance level (further details in Text S3 in the Supplementary Material). The following sections provide a detailed description of the different methodological steps.

### 3.1 Streamflow extremes characterization

We detect droughts, floods, and drought-to-flood transitions using the method proposed by Götte and Brunner (2024). This approach identifies periods of negative streamflow anomalies (i.e., droughts) using a daily varying threshold based on a 30-day rolling percentile of the daily streamflow data, while high streamflow events (i.e., floods) are identified using a fixed threshold based on a percentile of the annual maximum streamflow values. We further require that all drought events have a minimum

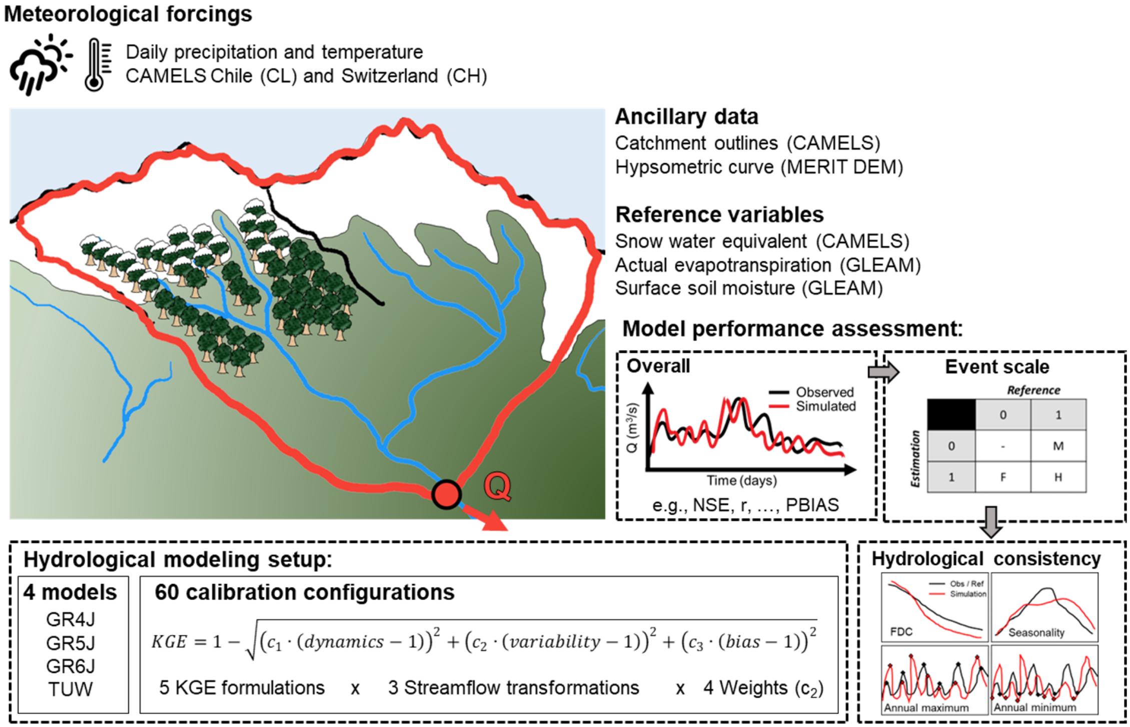


**Figure 2.** Overview of the methodological approach. See text for details

duration of 30 days, and we merge droughts separated by 15 days or less between them (Van Loon and Van Lanen, 2012; Fleig et al., 2006; Tallaksen et al., 1997) to limit the detection of minor events.

Rapid (within 14 days) and seasonal (within 90 days) transitions are defined based on the number of days between the end of the drought to the onset of the flood, following Götte and Brunner (2024). The thresholds for droughts (30th percentile of the smoothed daily flow) and floods (40th percentile of the annual maxima series) were set to ensure roughly one streamflow extreme event of each type (i.e., drought and flood) per year on average for each catchment. This target was set in order to identify a statistically representative number of extreme events, comparable to the sample size that would be obtained by the commonly used annual maximum approach (e.g., Meylan et al., 2012). Considering this definition and the thresholds adopted, we identified one event every four years on average for each catchment. To identify the thresholds that met our criteria, we tested different values (see Figure S1 in the Supplementary Material). Figure 3 illustrates the detection of droughts and floods based on the approach adopted for two recorded transitions in two catchments within the study domain.

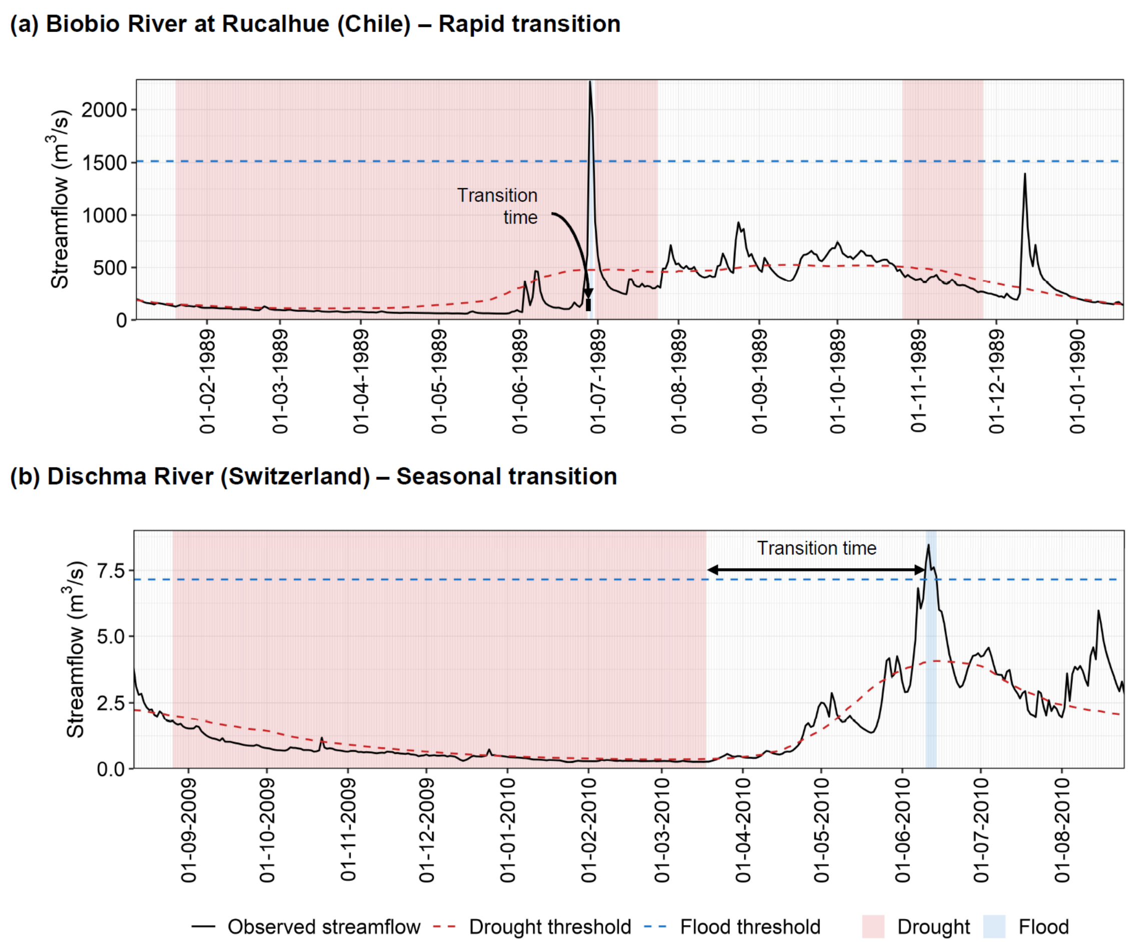


**Figure 3.** Example of the characterization of streamflow extremes and their transitions for two catchments within the study domain. (a) Biobio River at Rucalhue in Chile, and (b) Dischma River in Switzerland.

## 3.2 Modeling approach

### 3.2.1 Hydrological models

We use four conceptual bucket-style rainfall-runoff hydrological models: GR4J (Perrin et al., 2003), GR5J (Le Moine, 2008; Pushpalatha et al., 2011), GR6J (Pushpalatha et al., 2011), all coupled to the snow accumulation-ablation module CemaNeige (Valéry et al., 2014a, b), and TUWmodel (Parajka et al., 2007), which is based on the HBV model (Bergström and Forsman, 1973). All models have been widely used within the hydrological community during the last decades (Seibert and Bergström, 2022). GR4J, GR5J, and GR6J (with 6, 7, and 8 parameters coupled with CemaNeige, respectively; see Table S1 in the Supplementary Material) were chosen to explore how slight changes in model structure affect simulated streamflow extremes, and the TUW model (with 15 parameters; see Table S2 in the Supplementary Material) was selected to explore how more complex models, in terms of the snow routine and the representation of the processes occurring in the production storage, simulate these phenomena. To avoid overcompensating for biases in the meteorological forcings, which can be seen in the catchments placed above the water limit or below the energy limit - pointed out in Figure 1c -, two parameters were included in the calibration process in addition to the original setup for each hydrological model. Thus, a multiplicative parameter for precipitation (dP) and an additive parameter for temperature (dT) were included to adjust systematic biases in precipitation and temperature.

The GR4J, GR5J, and GR6J models - hereafter referred to as GRXJ for simplicity - share the same genealogy, meaning that they are based on the same core structure. These models can be coupled to the snow module CemaNeige, which partitions precipitation into liquid and solid and simulates snow accumulation and melt (rainfall and snowmelt enter the GRXJ structures). The basic structure of the GRXJ family corresponds to the GR4J model, which includes a parameter for production storage capacity, representing surface processes, and a parameter for routing storage capacity, representing subsurface processes. Additionally, GR4J includes an intercatchment exchange parameter and a unit hydrograph parameter that represents the delay between precipitation and streamflow. GR5J adds an additional parameter to the GR4J structure to improve the intercatchment exchange function, while GR6J includes a parameter for exponential storage in parallel to the routing storage included in GR4J and GR5J to improve the representation of groundwater processes (i.e., slow runoff).

The TUW model consists of a snow, soil, groundwater (subsurface flow), and a routing routine, similar to the HBV model (Bergström and Forsman, 1973). One of the major differences between the HBV and TUW models is in the snow routine. The TUW model does not allow for meltwater or rainfall to be retained within the snowpack, nor does it account for the refreezing of liquid water. The snow routine partitions between liquid and solid precipitation and estimates snow accumulation and melt. Rainfall and snowmelt enter the soil routine, where actual evaporation, soil moisture, and recharge are estimated. Then, the recharge flow goes to the groundwater routine, represented by two storages that produce surface runoff and quick flow (upper), and baseflow (lower). The sum of these flows is delayed in the routing routine using a triangular transfer function. Unlike the GRXJ models, which follow a water balance approach to characterize the production storage, TUW estimates evapotranspiration and recharge based on an explicit conceptualization of soil moisture content.

While both CemaNeige and the snow routine implemented in the TUW model follow a degree-day factor approach, there are differences in (i) the characterization of the precipitation phase (TUW allows the existence of a mixed partition between rain and snow), (ii) the conditions for snowmelt (free parameter in the TUW model and set as 0°C for CemaNeige), and (iii) the presence (or absence) of a parameter to correct for snowfall undercatch (not available in CemaNeige). These differences also explain the number of parameters that each of the snow routines has (two and five for CemaNeige and the snow routine in the TUW model, respectively).

Despite their structural differences and conceptualizations (for further details refer to Astagneau et al., 2021b), these models provide simplified representations of some hydrological states, fluxes, and processes at the catchment scale using precipitation (P), mean temperature (T), and potential evapotranspiration (PET) at daily time steps as inputs. To estimate PET, we use the approach proposed by Oudin et al. (2005), which is based on temperature and requires latitude and day of the year as a proxy for extraterrestrial radiation. Additionally, as the snow module CemaNeige can be configured in a semi-distributed way, discretizing each catchment into equal-area elevation bands based on the hypsometric curve, we considered 10 elevation bands for all evaluated model structures. To make simulations comparable across model structures, precipitation and temperature for the TUW model were extrapolated following the approach implemented in the GRXJ models through 10 elevation bands, based on the orographic gradients defined by Valéry et al. (2010).

#### 3.2.2 Calibration strategy

The parameters of each model structure, as well as the forcing adjustment parameters introduced, were calibrated using daily streamflow records and the Shuffled Complex Evolution global optimization algorithm (SCE-UA; Duan et al., 1992) over the period 2000-2020 (details on the convergence of the optimization algorithm and its configuration in Text S1 in the Supplementary Material). This calibration period was defined to capture the current hydroclimatic conditions in the modeling setup. Considering the temperature adjustment parameter, potential evapotranspiration was recalculated in each iteration during calibration to ensure consistency between those variables. Different objective functions based on the KGE configuration were used to calibrate each model. In its most general form, the KGE (Eq. (1)) compares simulations to a reference based on three components, i.e. dynamics (e.g., correlation), variability (e.g., standard deviation), and bias (e.g., mean). KGE values range from negative infinity to one, which is the optimum. How each component is defined depends on which KGE formulation is used. To the best of our knowledge, there exist five such formulations in the literature (Gupta et al., 2009; Kling et al., 2012; Pool et al., 2018; Tang et al., 2021; Pizarro and Jorquera, 2024, more details in Table 1). Additionally, different scaling factors or weights (i.e., $c_1$, $c_2$, and $c_3$ in Eq. (1)) can be used to put more emphasis to some of the components of the KGE as well as different streamflow transformations to give more weight to specific parts of the flow distribution (e.g., Thirel et al., 2024; Mizukami et al., 2019). To emphasize low flows, for example, flow can be transformed to the inverse of streamflow (i.e., 1/Q; e.g., Garcia et al., 2017; Wu et al., 2025). Further, linear combinations of the KGE applied to flows without and with transformation (i.e., Q and 1/Q, respectively) have been presented as a useful objective function to find a good compromise between high- and low-flows (e.g., Araya et al., 2023; Knoben et al., 2020; Muñoz-Castro et al., 2023).

$$\mathrm{KGE} = 1 - \sqrt{(c_1 \cdot (\mathrm{dynamics} - 1))^2 + (c_2 \cdot (\mathrm{variability} - 1))^2 + (c_3 \cdot (\mathrm{bias} - 1))^2} \qquad (1)$$

For each hydrological model and catchment, 60 different objective functions (OF) are implemented based on the possible combinations of the following methodological choices: (i) 5 KGE formulations (Table 1), (ii) 3 streamflow transformation cases (High, Low, High-Low), and (iii) 4 weights applied to the variability term (i.e., in Eq. (1), $c_2$ = 1, 2, 4, 8). For the low-flow transformation (Low; i.e., using 1/Q), a constant equal to 1% of the mean streamflow is added to the series to avoid zero-flow problems following recommendations from previous studies (e.g., Pushpalatha et al., 2012; Garcia et al., 2017; Knoben et al., 2020). To facilitate the notation associated with the streamflow transformations tested here, we will refer to the case 'Hi' (High) when a certain formulation of KGE is applied to untransformed streamflow (i.e., Q), while 'Lo' (Low) will refer to the case where a low-flow transformation is applied (i.e., 1/Q). We will refer to the linear combination of both cases (i.e., 0.5*Hi + 0.5*Lo) as 'HiLo'.

**Table 1.** Summary of KGE formulations. In each formulation, the term dynamics stands for the representation of the temporal evolution of the target variable, while the terms variability and bias aim to characterize its distribution.

| KGE formulation | Components | Description | Reference |
| --- | --- | --- | --- |
| Original (KGE) | **Dynamics:** Pearson correlation coefficient.<br>**Variability:** Ratio between the standard deviation of the simulated and observed values.<br>**Bias:** Ratio between the mean of the simulated and observed values. | Meta-objective function, oriented to quantify the Euclidean distance between the absolute error associated with each component. Proposed to overcome the problems associated with NSE (e.g., observed mean as baseline, formulation, which could lead to large volume balance errors or favor models/parameters sets that underestimate the observed variability). | Gupta et al. (2009) |
| Modified (KGE_mod1) | **Dynamics:** Pearson correlation coefficient.<br>**Variability:** Ratio between the coefficient of variation of the simulated and observed values.<br>**Bias:** Ratio between the mean of the simulated and observed values. | Modification in the variability component defined in the original formulation (i.e., standard deviation ratio) aimed to ensure that the bias and variability ratios are not cross-correlated. | Kling et al. (2012) |
| Non-parametric (KGE_np) | **Dynamics:** Spearman's rank correlation coefficient.<br>**Variability:** Error between all ranked simulated and observed values (i.e., flow duration curve) normalized to remove the volume information and keep only the distribution signal.<br>**Bias:** Ratio between the mean of the simulated and observed values. | Reformulation of the variability and correlation terms in a non-parametric way to address the implicit assumptions of linearity and normality of the data in the original formulation. | Pool et al. (2018) |
| Modified v2 (KGE_mod2) | **Dynamics:** Pearson correlation coefficient.<br>**Variability:** Ratio between the standard deviation of the simulated and observed values.<br>**Bias:** Ratio between the mean of the simulated minus the observed values and the standard deviation of the observed values. | Modification in the bias component defined in the original formulation aimed to avoid anomalously negative values when the mean value is close to zero. | Tang et al. (2021) |
| K-Moments (KGE_km) | **Dynamics:** Pearson correlation coefficient.<br>**Variability:** Ratio between the coefficient of variation of the simulated and observed values defined from unbiased estimators of non-central K-moments (alternative formulation for the second moment).<br>**Bias:** Ratio between the mean of the simulated and observed values. | Modification in the variability component defined in the original formulation oriented to make it less sensitive to outliers and non-normal distributions. | Pizarro and Jorquera (2024) |

## 3.3 Model performance assessment

### 3.3.1 Overall performance and hydrological consistency

To assess our results independently of the KGE configuration across the different calibration experiments tested here, we quantified the overall model performance using the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), which has become one of the most popular metrics within the hydrological community in the last decades (Melsen et al., 2025). We followed a traditional split-sample test approach (Klemeš, 1986; Beven, 2025) to assess the model performance over two time periods defined as (i) calibration (2000-2020), and (ii) evaluation (1985-1999).

We compute the NSE for different variables, including high- and low-flows (i.e., Q and 1/Q), snow water equivalent (SWE), soil moisture (SM), and actual evapotranspiration (ET). To minimize the influence of biases in ET and SM estimates on the estimates of model performance - due, for example, to catchment-scale averaging (e.g., Rouholahnejad Freund et al., 2020) or uncertainties related to the GLEAM algorithm (e.g., Jahromi et al., 2022) - we only considered the signal (i.e., timing and variability) associated with these variables rather than their absolute values. Complementarily, we estimate the bias associated with some hydrological signatures derived from the daily streamflow time series (e.g., mean, variance, Q1, Q99). To focus our study on assessing the ability of hydrological models to capture streamflow extreme events and their transitions, these results are presented in Text S2 in the Supplementary Material.

### 3.3.2 Detection of streamflow extreme events

To assess the model's capability to detect streamflow extremes and their transitions, we use the Critical Success Index (CSI; Eq. (2)), which is formulated based on the number of hits (H; events identified both in the reference/observation and the simulation), misses (M; events only identified in the reference/observation), and false alarm events (F; events identified only in the simulation). The CSI values vary between zero and one, with one being the optimum. We define hits as simulated events overlapping at least 50% with their observed counterparts. Additionally, for the detection analysis, a tolerance window of 30 and 5 days is defined before the onset and after the end of an observed drought and flood event, respectively. This window allows for considering the differences of gridded meteorological products with reality and how these can affect the timing of the simulated events (i.e., early or late compared to the observed streamflows). In short, we aim to evaluate the models' ability to capture streamflow extremes and their transitions rather than their characteristics (which may be even more restrictive). Therefore, we do not analyse the performance of the model in representing the specific characteristics of each event (e.g., cumulative deficit during the drought period, flow peak, etc.).

$$CSI = H/(H + M + F) \quad (2)$$

## 3.4 Assessment of the relative importance of modeling decisions

To assess the relative importance of modeling decisions on the detection of streamflow extremes and their transitions, we conduct an analysis of variance (ANOVA; Fisher, 1992; Kaufmann and Schering, 2014). The ANOVA enables us to examine the relationship between different modeling decisions (e.g. choice of structure and different decisions related to calibration) and quantify their relative importance in explaining the total variance in the target variable (e.g., CSI). Thus, by dividing the total variance into different groups, genuine sources of variation that are not explained by chance can be identified. We assume that the total variance (TV) in the target variable can be mainly explained by the differences between catchments (CT), hydrological models (HM), KGE formulations (KGEf), streamflow transformations (QTR), and KGE component weights (W). If, for example, weights do not have a significant impact on the detection of streamflow extremes, we would expect a low value for the term "W", that is a lower relative importance (i.e., W/TV) for explaining the total variance with respect to other decisions. Based on this conceptualization and considering a residual term (RS) that groups all the interactions between decisions and the variance that we cannot explain from them, the ANOVA can be expressed as follows:

$$TV = CT + HM + KGEf + QTR + W + RS \quad (3)$$

### 3.5 Identification of important processes in simulating drought-to-flood transitions

To identify the most important processes in simulating drought-to-flood transitions, we ask what explains the accurate detection of events. To address this question, we analyze the relative importance of each model parameter in estimating the CSI through an ANOVA test applied per catchment. This analysis, expressed by Eq. 4, considers the 60 alternative configurations (i.e., parameter sets) available per model and catchment and uses the total variance explained (TV) by each parameter ($\theta_i$; where i = 1, ..., Np, and Np is the number of parameters) as a proxy for the importance of the associated variable/process and a residual term (RS). The approach used to analyze the relative importance of the parameters explaining the variance of the CSI may have problems if the parameters do not show enough variation between the different configurations. However, despite the similarities in the configurations used for calibration, almost all the parameters show high variability among the calibrated parameter sets per catchment (see Figure S12 in Supplementary Material).

$$\mathrm{TV} = \sum_{i=1}^{Np} \theta_i + \mathrm{RS} \tag{4}$$

## 4 Results

The results presented here are based on the simulations with the HiLo (i.e., 0.5*KGE(Q) + 0.5*KGE(1/Q)) configuration, unless specific cases where all 60 configurations per catchment were used (e.g., ANOVA tests). This is considering that our results are consistent with other studies, which have shown that the use of such approaches enables a good compromise in simulating both low and high flows (e.g., Garcia et al., 2017; Thirel et al., 2024; Lema et al., 2025). The results for the alternative streamflow transformations are presented in the Supplementary Material.

### 4.1 Suitability of KGE for calibrating models aimed at simulating drought-to-flood transitions

To illustrate how closely the general model performance described by the KGE is linked to the capability of detecting extreme events, we compare the objective function value retrieved for one of our calibration configurations - the original KGE formulation configured with unweighted HiLo (i.e., $c_2$ = 1 and HiLo = 0.5*KGE(Q) + 0.5*KGE(1/Q)), which is later used as a reference - with the performance in detecting droughts, floods and their transitions based on the CSI (Figure 4). Our comparison clearly shows that model performance varies across catchments and model structures for both the KGE and CSI. While the overall performance described by the KGE can potentially be used as a proxy for a model's performance in capturing droughts for some catchments (e.g., points close to the optimal values for both KGE and CSI, i.e., 1, and CSI ranges from 0.23 to 0.79 for GR4J and from 0.21 to 0.74 for TUW), it is not generalizable to floods and transitions or to all the models tested here. Therefore, a high KGE does not necessarily imply a high CSI for these two types of events.

While KGE is not necessarily a good proxy for how well a model captures extreme events (especially floods and transitions), some specific KGE formulations might be better suited for this task than others. Next, we evaluate to what extent different

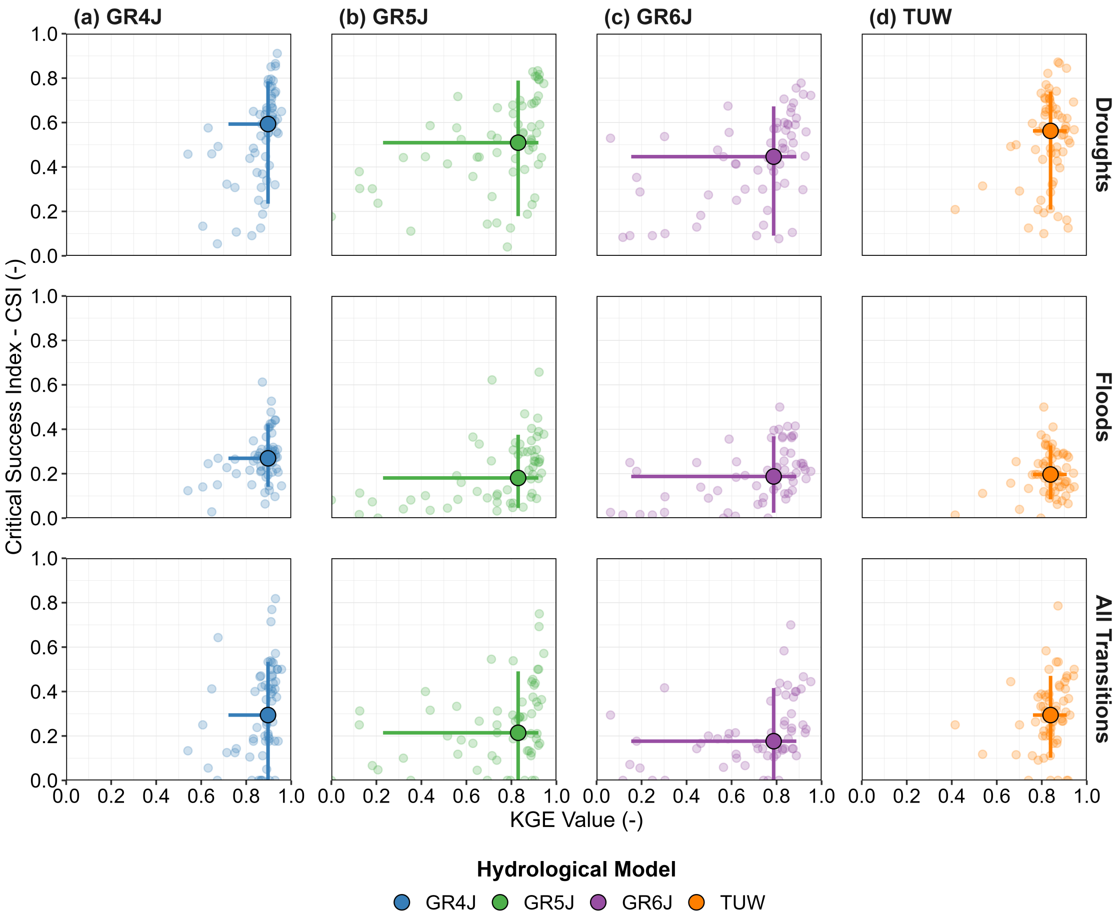


**Figure 4.** Comparison between the Kling-Gupta Efficiency (KGE) for the calibration period and the Critical Success Index (CSI) for droughts, floods, and transitions, based on the simulations with the models (a) GR4J, (b) GR5J, (c) GR6J, and (d) TUW calibrated with the unweighted original KGE formulation as the objective function. The dispersion bars are associated with the 10th and 90th percentiles across catchments, while the central shape is associated with the 50th percentile. Circles with transparency show results for each catchment. For both KGE and CSI, the optimal value is 1.

adjustments in the 'basic' configuration used for the analysis presented above can (or cannot) improve the performance in detecting streamflow extreme events and, particularly, drought-to-flood transitions.

### 4.2 Impacts of KGE configurations on drought-to-flood transition simulations

To assess the added value of the application of weights to the variability term of the KGE as well as the use of different KGE formulations for detecting independent extreme events and their transitions, Figure 5 shows - for the GR4J model as an example - the differences in CSI between the unweighted original KGE (reference) and alternative cases (e.g., weights

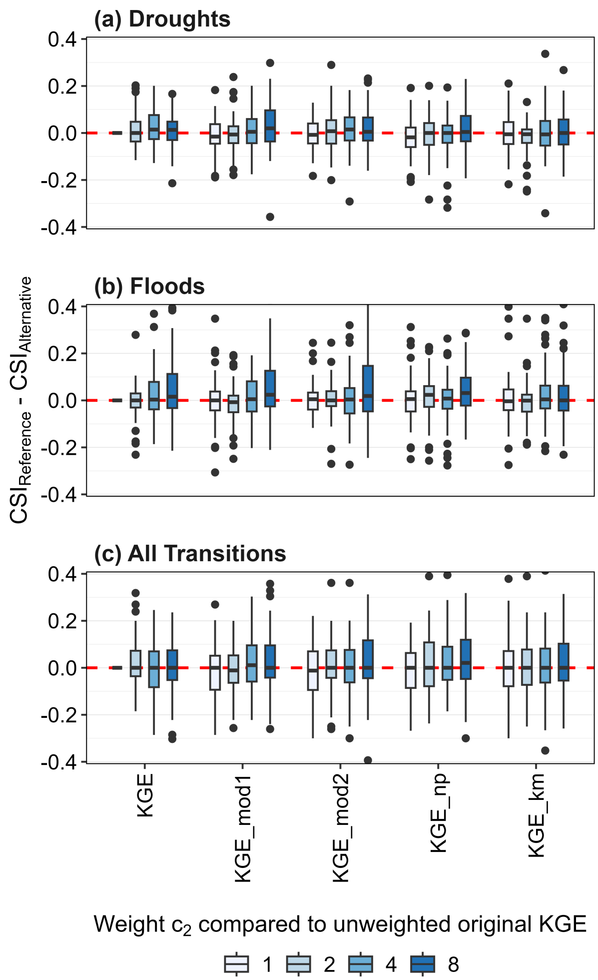


**Figure 5.** Difference in the CSI for GR4J simulations using model calibrations with no weights and the original KGE (reference) versus different weights and KGE formulations (alternative) for (a) droughts, (b) floods, and (c) transitions. Values above (below) 0 indicate better (worse) performance of the reference compared to the alternative. Each boxplot displays the information of 63 values (i.e., one per catchment).

and/or KGE formulations).These results highlight that, in the context of a large-sample study, weighting the variability term of the KGE does not consistently enhance model performance in detecting streamflow extremes and their transitions (median difference is centered around 0 in both cases) and may even be detrimental.In short, the use of weights and the choice of the KGE formulation do not play a dominant role in the overall performance of the model over the study domain, and its usefulness will depend on the characteristics of the study domain and the model structure (see Figures S6 in the Supplementary Material). These findings are consistent across other model structures tested (Figures S4-S11 in Supplementary material).

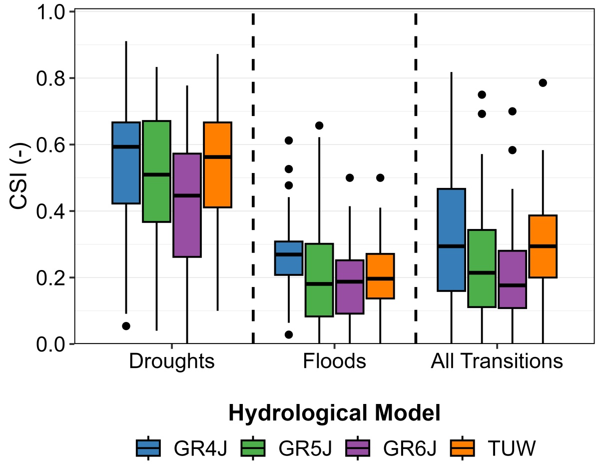


**Figure 6.** Critical Success Index (CSI) for (a) droughts, (b) floods, and (c) drought-to-flood transitions, based on the simulations with GR4J, GR5J, GR6J, and TUW (different colors) calibrated with the unweighted HiLo original KGE formulation as the objective function. Each boxplot displays the information of 63 values (i.e., one per catchment).

### 4.3 Importance of model structure

Our results show that the detection of droughts is typically more reliable than flood and transition detection (Figure 6). However, there are substantial differences in the detection rate depending on the hydrological model structure used for simulating extreme events. GR4J yields the overall best performance regardless of the KGE formulation chosen for calibration (median CSI values across KGE formulations around 0.58, 0.26, and 0.31 for droughts, floods, and transitions, respectively). The performance of the GR5J and GR6J models decreases compared to GR4J (changes between 0.05 and 0.13 depending on the model and type of extreme event). This suggests that increasing model complexity decreases rather than increases model performance in detecting streamflow extreme events. However, these decreases in performance cannot be directly attributed to increases in the number of parameters because the TUW model shows comparable results to GR4J despite its structure (median CSI values across KGE formulations around 0.56, 0.19, and 0.28 for droughts, floods, and transitions, respectively). If model structures are compared for Switzerland and Chile (see Figure S7 in Supplementary Material), the same conclusions can be drawn in terms of model performance, with comparable results between GR4J and TUW and lower performance of the GR5J and GR6J models. However, the detection of extreme events is more challenging in catchments located in Chile compared to those located in Switzerland, with differences in the median CSI between countries being around 0.23, 0.02, and 0.13 for droughts, floods, and drought-to-flood transitions, respectively. In summary, the GR4J and TUW models seem to be the models best suited for simulating droughts and floods among the models considered in this study.

Different model structures can result in similar streamflow simulations even though they represent hydrological fluxes and states in different ways. To illustrate this, we compare simulated fluxes obtained for an observed seasonal drought-to-flood transition in the Dischma river in Switzerland across the four hydrological models (Figure 7). While three out of four models capture the transition event successfully (GR6J fails in capturing its timing) and show similar temporal patterns of ET, snowmelt, and SWE, the contribution of baseflow (presented as a percentage of total runoff) and soil moisture vary strongly among them. Consequently, the analysis of the drivers associated with such transition events will vary depending on which model structure is analyzed. Although there is a high agreement between the models in terms of the detection of the event in this example case (i.e., 3 out of 4), this is not necessarily the case for all events and catchments (Figure 4).

### 4.4 Relative importance of different modeling decisions

In agreement with our earlier findings (Figure 6), the results of the ANOVA show that the most important modeling decision is the choice of a suitable model structure, followed by the choice of the streamflow transformation, and the differences between catchments (Figure 8). In contrast, the choices of KGE formulation and weights do not have a strong impact on the performance in simulating streamflow extremes. The relative importance of the methodological choices is similar when analyzing other categorical indices, such as the probability of detection, false alarm ratio, and frequency of bias (see Figure S8 in the Supplementary Material). For rapid transitions, the difference between catchments is more important for explaining the CSI values than it is for seasonal transitions. This difference indicates that the detection of rapid transition events depends even more strongly on catchment attributes (e.g., mean elevation, streamflow regime, etc.).

### 4.5 Model performance depends on catchment characteristics

We explore the relationship between model performance and catchment characteristics using Spearman's rank correlation coefficient. To this end, we focus on the CSI obtained for the different types of extreme events of interest (droughts, floods, and transitions) generated with the GR4J and TUW models calibrated with the unweighted HiLo original KGE formulation (Figure 9).

Our results show that a model's capability in simulating extreme hydrological events and their transitions depends on catchment characteristics (Figure 9). Drought-to-flood transitions are more difficult to capture in semi-arid (negative correlation between aridity index and CSI), high-mountain (negative correlation between mean elevation and CSI), and flashy (negative correlation between the slope of the flow duration curve and CSI) catchments than in humid low-elevation catchments with high streamflow elasticity to precipitation. This result is generalizable to the other models and the different KGE formulations tested (see Figure S9 in the Supplementary Material).

### 4.6 Linking model performance to hydrological processes during streamflow extremes

We conduct an ANOVA test to analyze the relative importance of different model parameters in detecting streamflow extremes and their transitions (Figure 10; the extended version with rapid and seasonal transitions is presented in Figure S13 in the Supplementary Material). Here, we show that some model parameters are relatively more important than others (e.g., X4 for floods in GRXJ models), but that the relative importance of a given parameter can vary substantially across catchments. All the hydrological models show a high importance of the parameters aimed to adjust the forcings (i.e., dP and dT for all the models

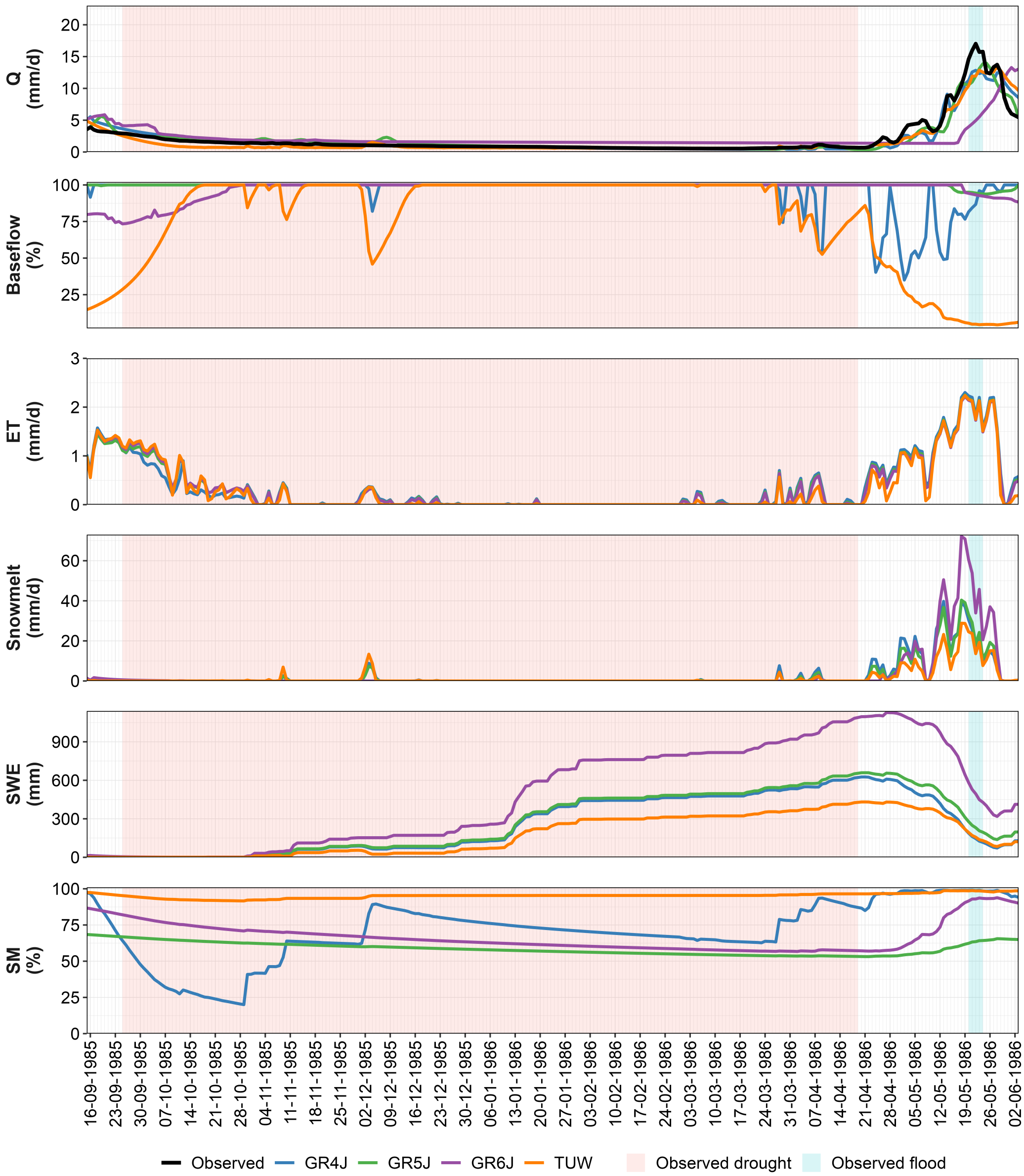


**Figure 7.** Example of how different hydrological fluxes and states are simulated for an observed drought-to-flood transition in the Dischma river (Switzerland) with the GR4J, GR5J, GR6J, and TUW hydrological models calibrated with the unweighted HiLo original KGE formulation.

as well as SCF in TUW model, which seeks to correct for the snow undercatch), highlighting the need for adequate forcing to improve the estimation of extreme hydrological events. For the GRXJ models, X3 (routing store capacity) and X4 (unit hydrograph time constant) are more important in the simulation of low and high flow compared to the rest of the parameters,

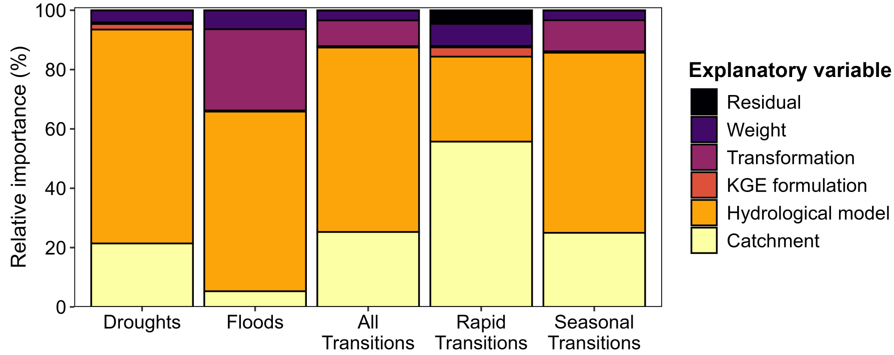


**Figure 8.** Results of the analysis of variance (ANOVA) applied to the Critical Success Index (CSI) for droughts, floods, all drought-to-flood transitions (i.e., rapid and seasonal), rapid transitions (<14 days), and seasonal transitions (<90 days).

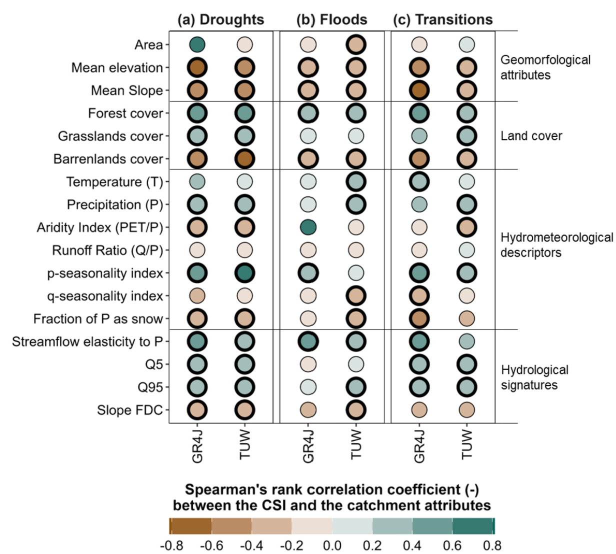


**Figure 9.** Spearman's rank correlation coefficient between different catchment attributes and the CSI for (a) droughts, (b) floods, and (c) drought-to-flood transitions, based on the simulations with GR4J and TUW calibrated using the unweighted HiLo original KGE formulation as the objective function. The circles with thick outlines indicate statistically significant correlation coefficients at a the 5% significance level.

which is accentuated even more when more complexity is added to the base structure (i.e., GR6J). In the TUW model, which has more parameters than the GRXJ structures, the relative importance of each parameter is more uniform, and their relative importance is low, except for the parameter k0 (storage coefficient for very fast response), which becomes more important for flood detection.

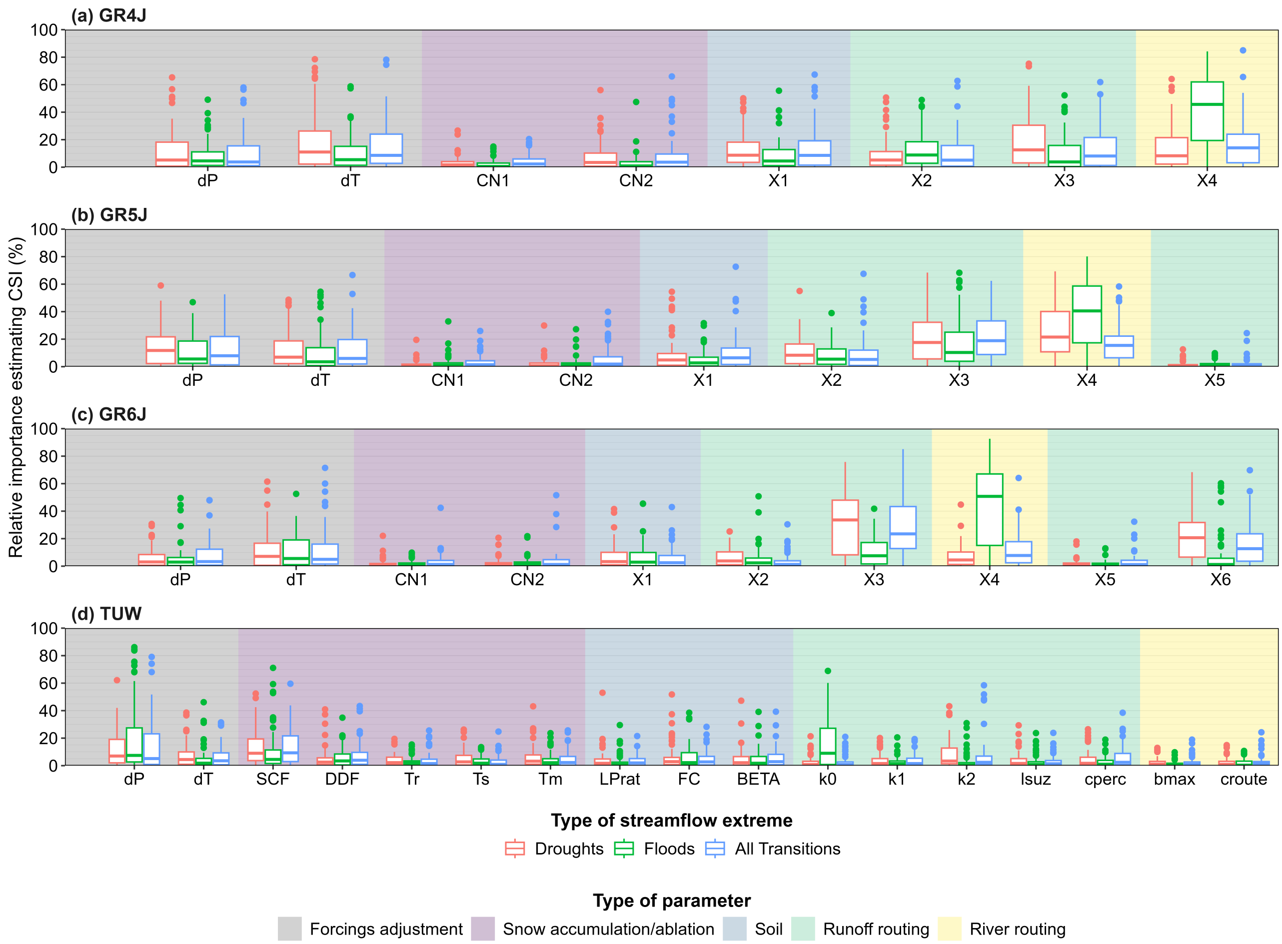


**Figure 10.** Relative importance of parameters for explaining the Critical Success Index (CSI) for models (a) GR4J, (b) GR5J, (c) GR6J, and (d) TUW based on the results of an analysis of variance (ANOVA).

# 5 Discussion

## 5.1 Simulation of compounding streamflow extreme events

We find that the hydrological models tested are better at detecting droughts (median CSI across catchments and KGE formulations: 0.45-0.58 depending on the model) than floods (median CSI across catchments and KGE formulations: 0.15-0.26 depending on the model), and their performance in detecting drought-to-flood transitions is closely related to the performance in detecting floods (Figure 4). This difference in drought and flood simulation performance can be attributed to the different timescales associated with these two types of extreme events: while droughts vary in duration from months to years (or decades), floods develop, and may also subside, in a matter of hours or days. This is consistent with the poor performance of all the models tested in capturing rapid transitions (i.e., occurring within 14 days; median CSI equal to zero when rapid and seasonal transitions are analyzed separately; not shown). Moreover, only in 13 basins - based on different configurations - we obtained CSI values greater than zero for rapid transitions (Table S4 in Supplementary Material). Our analyses highlight that these fast processes are rather difficult to capture in conceptual rainfall-runoff models.

### 5.2 Good general performance does not imply that extremes are well detected

Our results highlight that a good general model performance in terms of KGE does not necessarily imply a good performance in detecting streamflow extremes. Even models with KGE$>0.6$ struggle to capture extreme events, in particular floods and transitions from droughts to floods (Figure 4). These findings are aligned with previous studies discussing the potential of KGE to represent high-flow values or capture flashy dynamics (e.g., Astagneau et al., 2022; Brunner et al., 2021; Mathevet et al., 2020). For instance, Astagneau et al. (2021a) demonstrated that the relationship between KGE values and a model's capability of simulating summer floods is weak. Similarly, Bruno et al. (2024) showed that, during extreme low-flow conditions, model performance is usually lower than during normal flow conditions. Spieler and Schütze (2024) showed that the KGE lacks the capacity to provide information about detailed processes, leading to gaps between model accuracy (i.e., how well a model matches simulations with observations) and adequacy (i.e., how well a model captures key processes and behaviors of the observed system). These findings suggest that the traditional evaluation of hydrological models through goodness-of-fit metrics such as KGE or NSE must be accompanied by an explicit examination of their capability to simulate and detect streamflow extreme events, e.g. by using metrics such as the CSI.

### 5.3 The importance of different modeling decisions for simulating streamflow extremes and their transitions

Our results show that model structure is the most important modeling decision for capturing extreme events and their transitions (Figure 8), which is consistent with previous studies focused on the independent analysis of extreme events (e.g., Alexander et al., 2023; Melsen and Guse, 2019; van Kempen et al., 2021). Among the structures tested, the GR4J model provided the best performance both for the simulation of independent extreme events and for transitions (Figure 6). The TUW model shows similar results but with a lower flood detection performance (median CSI between KGE formulations of 0.19 compared to 0.26 for GR4J). These deficiencies in flood simulation performance translate to deficiencies in capturing drought-to-flood transitions (Figure 4 and Figure 6). The lack of an explicit structural component that allows for the simulation of floods that occur under dry conditions with low soil moisture could explain the poor performance associated with this type of compound event. Astagneau et al. (2022) highlighted that conditioning the storages and fluxes of a lumped conceptual hourly-timestep model on rainfall intensities could benefit model performance in catchments with a fast response to precipitation (i.e., flashy-catchments). For droughts, van Kempen et al. (2021) have shown that the magnitudes of the low-flow events are significantly affected by alterations in the architecture of the upper and lower storages, which is consistent with the changes in performance among the GRXJ models, where small structural modifications lead to important changes in the detection of these events.

We demonstrated that the capability to identify streamflow extreme events and their transitions in simulations is model-dependent. However, the change in performance in detecting extreme events does not necessarily depend on the number of parameters or model complexity. We obtained similar performances between GR4J and TUW despite their structural differences (Figure 6), unlike GR5J and GR6J, whose performance declines in comparison with GR4J. Several studies have highlighted that including a more detailed representation of hydrological processes in models does not necessarily imply better performance (e.g., Orth et al., 2015; Valéry et al., 2014a). This is because more realistic representations require more detailed information to characterize the system of interest (e.g., land cover maps, distributed forcings, high-resolution digital terrain model, soil properties), which is not always available. Recently, Santos et al. (2025) found that models with varying complexity can lead to similar robustness issues, stressing the need to improve strategies for diagnosing the suitability of model structures to improve the understanding of specific processes (e.g., Spieler and Schütze, 2024; Knoben et al., 2020).

The results presented here show that the choice of objective function is relatively less important compared to the choice of model structure (Figure 8). However, model performance can be optimized both in terms of general performance (NSE) and the representation of extreme events (CSI) under (a) the application of equal weights to all components of the KGE (Figure 5) and (b) the application of a streamflow transformation that focuses on both high and low flows (Figure S6 and S7 in the Supplementary Material). Our comparison also highlights that the potential benefit from adjusting these choices (e.g., using other weights or other transformations) varies widely between catchments (Figure 5). This is in line with the findings of Mizukami et al. (2019), who found that the influence of weights on model performance depends on model structure and catchment characteristics. While none of the tested modifications in the objective function consistently improve the simulation of streamflow extremes across all catchments in the study domain, some of the alternative KGE formulations could improve the simulation of certain variables.

Given the relative importance shown by the forcing adjustment parameters (i.e., dP and dT for all the models as well as SCF in TUW model, which seeks to correct the snow undercatch; Figure 10), the meteorological forcings can also have a major impact for detecting streamflow extremes and their transitions. Several studies have shown that errors in meteorological forcing are a key challenge in hydrological modeling (e.g., Brunner, 2023; Döll et al., 2016) due to, e.g., their large influence on the simulation of snow processes (e.g., Tang et al., 2023; Günther et al., 2019), or significant impacts on the partitioning between evaporation and runoff (e.g., Nasonova et al., 2011). Here, we attempt to reduce this effect by (1) utilizing local meteorological products over global ones, based on the evidence that these may enhance hydrological modeling (e.g., Clerc-Schwarzenbach et al., 2024), and (2) incorporating adjustment factors to account for potential systematic biases associated with them (e.g., Hughes, 2024; Probst and Mauser, 2022). However, introducing forcing adjustment factors could compensate for some model deficiencies by modifying the inputs (e.g., Tang et al., 2023, 2025). This is somehow reflected by the high dispersion of forcing adjustment factors within each configuration (Figure S11 in the Supplementary Material). Therefore, an improvement in the spatiotemporal representation of precipitation and temperature, as well as of the potential interactions between these variables, could contribute to improved representations of compound streamflow extreme events in hydrological models.

### 5.4 Limitations and recommendations for future work

Our model calibration experiments focused on the simulation of extreme streamflow events, which required the choice of specific event definitions. Here, we defined hydrological droughts and floods using threshold-based approaches, and the thresholds were adjusted in a way to identify, on average, one event per year and catchment. Because this methodological choice does, to a certain degree, affect the outcomes of the comparison, we tested different thresholds for defining streamflow extreme events. The results of this sensitivity analysis indicate that using more flexible thresholds to define droughts (i.e., higher percentiles) can enhance the detection of these events, as more instances are identified, and they tend to be less severe compared to more restrictive thresholds. However, we did not find such an effect for floods and transitions, for which we obtained similar model performances regardless of the thresholds used (see Figure S2 in Supplementary Material). The improvement in drought detection when the threshold is relaxed can be explained by the fact that models generally struggle during more extreme hydrological drought periods (e.g., Bruno et al., 2024), which are relatively less frequent if the threshold is raised. Similar results are obtained when the overlap window used to identify the hits is modified (Figure S3). While our study shows that the choice of threshold does not substantially affect model performance in terms of transition events, the method used to define streamflow extreme events can have a major impact on the characteristics of the transition events identified.

To support our analysis, we tested four bucket-type hydrological models used within the hydrological modeling community (Addor and Melsen, 2019). Even though these models are at the lower end in terms of model complexity (Hrachowitz and Clark, 2017), and three of them share the same core structure, they allowed us to perform a comprehensive analysis of different model structures at a lower computational cost than when using models with more complex structures (e.g., Clark et al., 2017; Orth et al., 2015; Poncelet et al., 2017). Furthermore, previous studies have also shown that more complexity does not necessarily imply better performance (Figure 6; e.g., Li et al., 2015; Merz et al., 2022). These models have been calibrated based on daily streamflow records, assuming that the numerical convergence of the optimization algorithm ensures (to some extent) a successful calibration process (see Text S1 in the Supplementary Material). However, it is important to acknowledge that potential compensations for biases in meteorological forcings or model deficiencies can make the "optimal" parameter sets less identifiable (e.g., Clark and Vrugt, 2006; Vrugt et al., 2005; Beven, 2025). Here, we explore the (dis)agreement between the optimal parameters for each configuration (Figure S12 in the Supplementary Material), showing overall agreement indices of around 0.5 (i.e., the parameters have a range of variation of approximately 50% of the parameter space). This highlights the need to incorporate, for example, hydrological variables such as SWE or ET, to (i) complement model assessment, (ii) better define the parameter exploration range (Figure S10 in the Supplementary Material), and (iii) lead to parameter sets that ensure reliability and fidelity in representing hydrological processes.

Our results provide insights on possible avenues of future research that could benefit drought-to-flood transitions modeling, which include: (1) exploring the use of modular platforms and a multi-model ensemble approach to quantify model uncertainty and identify more suitable model structures (e.g., Saavedra et al., 2022); (2) improving our understanding of the role of the spatial variability of precipitation for accurate flood simulations (e.g., Macdonald et al., 2025; Astagneau et al., 2022); (3) assessing the benefits of model runs at a subdaily timestep (e.g., hourly); and (4) exploring alternative data-driven modeling approaches such as long short-term memory (LSTM) networks (e.g., Frame et al., 2022; Acuña Espinoza et al., 2025; Kratzert et al., 2018). Additionally, exploring relationships between the occurrence or characteristics (e.g., duration, severity) of this type of hydrological extreme events and some large-scale climate patterns (Garreaud et al., 2020; Marengo and Espinoza, 2016; Sun et al., 2016; De Luca et al., 2020) could improve their predictability.

## 6 Conclusions

We performed a modeling intercomparison study to (i) explore to what extent hydrological models can simulate drought-to-flood transitions and (ii) identify suitable modeling choices aimed at capturing these compound extreme events. For this intercomparison, we calibrated four conceptual bucket-type hydrological models (GR4J, GR5J, GR6J, and TUW) for 63 catchments in Chile and Switzerland using 60 different configurations of the Kling-Gupta Efficiency (KGE) as objective functions, based on five KGE formulations, four scaling factors, and three streamflow transformations. Based on the results of this intercomparison, we draw the following conclusions:

1. A satisfactory general model performance, as expressed by the KGE, does not guarantee a good performance in terms of detecting streamflow extremes and their transitions. While KGE can serve as a rough proxy for low-flow performance, it cannot for high-flows and drought-to-flood transitions. Consequently, assessments of the suitability of hydrological models for simulating extreme events and their transitions should be complemented with metrics describing extreme event detection performance such as the critical success index (CSI).

2. The most important modeling decision when it comes to simulating floods, droughts, and their transitions is the choice of a suitable model structure. Here, we demonstrate that the GR4J and TUW models have similar performance - with GR4J being slightly better at detecting floods and transitions - and adding model complexity by increasing the number of parameters does not necessarily improve the representation of extreme events.

3. In contrast, the choice of the objective function and its exact configuration are less important. The choice of a suitable streamflow transformation can improve the simulation of extreme events to a certain degree. Specifically, a joint focus on high and low flows by equally weighting the two streamflow transformations in the objective function (referred to as HiLo in our analysis) can improve model performance without compromising its ability to capture streamflow extremes. However, the choice of the exact KGE formulation and the use of weights for the variability term of the KGE do not substantially affect the simulation of extreme events and the direction of this effect depends on the catchment.

4. A model's performance in simulating streamflow extremes and transitions primarily depends on how well it captures streamflow timing rather than other hydrological signatures or variables such as evapotranspiration or snow water equivalent.

5. Drought-to-flood transitions are more difficult to capture in semi-arid, high-mountain, and flashy catchments than in humid low-elevation catchments.

This methodological intercomparison highlights that simulating streamflow extremes and their transitions is not a trivial modeling task and continued research is needed to improve model performance. The results of this intercomparison study suggest that time is best invested when focusing on improving model structures rather than calibration procedures. Specifically, hydrological model development should focus on improving the representation of processes and components associated with the temporal dynamics of discharge, such as routing or the soil response to intense snowmelt and rainfall. Additionally, the strong link between model performance and parameters aimed at correcting precipitation inputs suggests that the representation of extreme events could also be improved by investing in the quality of meteorological forcing datasets. Investments in improving the simulation of extreme events and their transitions are crucial because hydrological models can not only support process understanding related to compounding streamflow extremes, but also can be used to forecast such events at short time scales and to project future changes in the occurrence of drought-to-flood transitions. Such applications are critical to ensure society's preparedness for these types of hydrometeorological extreme events.

*Code and data availability.* The R-scripts and data used to produce the results shown in this paper - such as parameter sets used to generate the simulations and performance metrics - are publicly available through Zenodo (Muñoz-Castro et al., 2025, https://doi.org/10.5281/zenodo.17086604). CAMELS-CL (Alvarez-Garreton et al., 2018a) is available at PANGAEA (Alvarez-Garreton et al., 2018b) and https://camels.cr2.cl/ while CAMELS-CH (Höge et al., 2023a) at Zenodo (Höge et al., 2023b). The GLEAM3.8a dataset (Miralles et al., 2011) is available upon request at https://www.gleam.eu/.

*Author contributions.* EMC, BJA, and MIB conceptualized and designed the study. EMC conducted all the model simulations, analysis of results, and created the figures. All authors contributed to discussing the methodology and results and to reviewing and editing the manuscript.

*Competing interests.* One of the co-authors is a member of the editorial board of Hydrology and Earth System Sciences (HESS).

*Acknowledgements.* We thank the Swiss National Science Foundation for funding this project through grant 200021_214907. PM acknowledges the support from ANID-PIA Project AFB230001 (AMTC).

## References


Acuña Espinoza, E., Loritz, R., Kratzert, F., Klotz, D., Gauch, M., Álvarez Chaves, M., and Ehret, U.: Analyzing the generalization capabilities of a hybrid hydrological model for extrapolation to extreme events, Hydrology and Earth System Sciences, 29, 1277–1294, https://doi.org/10.5194/hess-29-1277-2025, publisher: Copernicus GmbH, 2025.

Addor, N. and Melsen, L. A.: Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models, Water Resources Research, 55, 378–390, https://doi.org/10.1029/2018WR022958, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR022958, 2019.

Alexander, A. A., Kumar, D. N., Knoben, W. J. M., and Clark, M. P.: Evaluating the parameter sensitivity and impact of hydrologic modeling decisions on flood simulations, Advances in Water Resources, 181, 104 560, https://doi.org/10.1016/j.advwatres.2023.104560, 2023.

Althoff, D. and Rodrigues, L. N.: Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment, Journal of Hydrology, 600, 126 674, https://doi.org/10.1016/j.jhydrol.2021.126674, 2021.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrology and Earth System Sciences, 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, publisher: Copernicus GmbH, 2018a.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: Links to files - The CAMELS-CL dataset, https://doi.org/10.1594/PANGAEA.894885, 2018b.

Alvarez-Garreton, C., Boisier, J. P., Garreaud, R., Seibert, J., and Vis, M.: Progressive water deficits during multiyear droughts in basins with long hydrological memory in Chile, Hydrology and Earth System Sciences, 25, 429–446, https://doi.org/10.5194/hess-25-429-2021, publisher: Copernicus GmbH, 2021.

Araya, D., Mendoza, P. A., Muñoz-Castro, E., and McPhee, J.: Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling, Hydrology and Earth System Sciences, 27, 4385–4408, https://doi.org/10.5194/hess-27-4385-2023, publisher: Copernicus GmbH, 2023.

Asadieh, B. and Krakauer, N. Y.: Global change in streamflow extremes under climate change over the 21st century, Hydrology and Earth System Sciences, 21, 5863–5874, https://doi.org/10.5194/hess-21-5863-2017, publisher: Copernicus GmbH, 2017.

Astagneau, P. C., Bourgin, F., Andréassian, V., and Perrin, C.: When does a parsimonious model fail to simulate floods? Learning from the seasonality of model bias, Hydrological Sciences Journal, 66, 1288–1305, https://doi.org/10.1080/02626667.2021.1923720, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2021.1923720, 2021a.

Astagneau, P. C., Thirel, G., Delaigue, O., Guillaume, J. H. A., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W., and Beven, K. J.: Technical note: Hydrology modelling R packages – a unified analysis of models and practicalities from a user perspective, Hydrology and Earth System Sciences, 25, 3937–3973, https://doi.org/10.5194/hess-25-3937-2021, publisher: Copernicus GmbH, 2021b.

Astagneau, P. C., Bourgin, F., Andréassian, V., and Perrin, C.: Catchment response to intense rainfall: Evaluating modelling hypotheses, Hydrological Processes, 36, e14 676, https://doi.org/10.1002/hyp.14676, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.14676, 2022.

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., Nauditt, A., Ribbe, L., and Thinh, N. X.: On the selection of precipitation products for the regionalisation of hydrological model parameters, Hydrology and Earth System Sciences, 25, 5805–5837, https://doi.org/10.5194/hess-25-5805-2021, publisher: Copernicus GmbH, 2021.

Banfi, F. and De Michele, C.: Compound flood hazard at Lake Como, Italy, is driven by temporal clustering of rainfall events, Communications Earth & Environment, 3, 1–10, https://doi.org/10.1038/s43247-022-00557-9, publisher: Nature Publishing Group, 2022.

Barendrecht, M. H., Matanó, A., Mendoza, H., Weesie, R., Rohse, M., Koehler, J., de Ruiter, M., Garcia, M., Mazzoleni, M., Aerts, J. C. J. H., Ward, P. J., Di Baldassarre, G., Day, R., and Van Loon, A. F.: Exploring drought-to-flood interactions and dynamics: A global case review, WIREs Water, n/a, e1726, https://doi.org/10.1002/wat2.1726, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wat2.1726, 2024.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model, Nordic Hydrol., 4, 147–170, https://doi.org/10.2166/nh.1973.0012, 1973.

Beven, K. J.: A short history of philosophies of hydrological model evaluation and hypothesis testing, WIREs Water, 12, e1761, https://doi.org/10.1002/wat2.1761, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wat2.1761, 2025.

Boisier, J. P.: CR2MET: A high-resolution precipitation and temperature dataset for the period 1960-2021 in continental Chile., https://doi.org/10.5281/zenodo.7529682, 2023.

Brunner, M. I.: Floods and droughts: a multivariate perspective, Hydrology and Earth System Sciences, 27, 2479–2497, https://doi.org/10.5194/hess-27-2479-2023, publisher: Copernicus GmbH, 2023.

Brunner, M. I. and Sikorska-Senoner, A. E.: Dependence of flood peaks and volumes in modeled discharge time series: Effect of different uncertainty sources, Journal of Hydrology, 572, 620–629, https://doi.org/10.1016/j.jhydrol.2019.03.024, 2019.

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., and Clark, M.: Spatial Dependence of Floods Shaped by Spatiotemporal Variations in Meteorological and Land-Surface Processes, Geophysical Research Letters, 47, e2020GL088 000, https://doi.org/10.1029/2020GL088000, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088000, 2020.

Brunner, M. I., Melsen, L. A., Wood, A. W., Rakovec, O., Mizukami, N., Knoben, W. J. M., and Clark, M. P.: Flood spatial coherence, triggers, and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated models, Hydrology and Earth System Sciences, 25, 105–119, https://doi.org/10.5194/hess-25-105-2021, publisher: Copernicus GmbH, 2021.

Brunner, M. I., Anderson, B., and Muñoz-Castro, E.: Meteorological and hydrological dry-to-wet transition events are only weakly related over European catchments, Environmental Research Letters, 20, 084 013, https://doi.org/10.1088/1748-9326/ade72c, publisher: IOP Publishing, 2025.

Bruno, G., Avanzi, F., Alfieri, L., Libertino, A., Gabellani, S., and Duethmann, D.: Hydrological model skills change with drought severity; insights from multi-variable evaluation, Journal of Hydrology, 634, 131 023, https://doi.org/10.1016/j.jhydrol.2024.131023, 2024.

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., and Goldscheider, N.: When best is the enemy of good – critical evaluation of performance criteria in hydrological models, Hydrology and Earth System Sciences, 27, 2397–2411, https://doi.org/10.5194/hess-27-2397-2023, publisher: Copernicus GmbH, 2023.

Clark, M. P. and Vrugt, J. A.: Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL025604, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005GL025604, 2006.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences, 21, 3427–3440, https://doi.org/10.5194/hess-21-3427-2017, publisher: Copernicus GmbH, 2017.

Clerc-Schwarzenbach, F., Selleri, G., Neri, M., Toth, E., van Meerveld, I., and Seibert, J.: Large-sample hydrology – a few camels or a whole caravan?, Hydrology and Earth System Sciences, 28, 4219–4237, https://doi.org/10.5194/hess-28-4219-2024, publisher: Copernicus GmbH, 2024.

Cortés, G. and Margulis, S.: Impacts of El Niño and La Niña on interannual snow accumulation in the Andes: Results from a high-resolution 31 year reanalysis, Geophysical Research Letters, 44, 6859–6867, https://doi.org/10.1002/2017GL073826, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL073826, 2017.

De Luca, P., Messori, G., Wilby, R. L., Mazzoleni, M., and Di Baldassarre, G.: Concurrent wet and dry hydrological extremes at the global scale, Earth System Dynamics, 11, 251–266, https://doi.org/10.5194/esd-11-251-2020, publisher: Copernicus GmbH, 2020.

Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaefli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, Water Resources Research, 56, e2019WR026 085, https://doi.org/10.1029/2019WR026085, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR026085, 2020.

Di Baldassarre, G., Martinez, F., Kalantari, Z., and Viglione, A.: Drought and flood in the Anthropocene: feedback mechanisms in reservoir operation, Earth System Dynamics, 8, 225–233, https://doi.org/10.5194/esd-8-225-2017, publisher: Copernicus GmbH, 2017.

Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., Damm, B., Lucia, A., Fauer, F. S., Nissen, K. M., Sieg, T., and Thieken, A. H.: More than heavy rain turning into fast-flowing water – a landscape perspective on the 2021 Eifel floods, Natural Hazards and Earth System Sciences, 22, 1845–1856, https://doi.org/10.5194/nhess-22-1845-2022, publisher: Copernicus GmbH, 2022.

Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D.: The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review, PLoS Currents, 5, ecurrents.dis.f4deb457904 936b07c09daa98ee8171a, https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a, 2013.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015–1031, https://doi.org/10.1029/91WR02985, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/91WR02985, 1992.

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

Fatichi, S., Rimkus, S., Burlando, P., Bordoy, R., and Molnar, P.: High-resolution distributed analysis of climate and anthropogenic changes on the hydrology of an Alpine catchment, Journal of Hydrology, 525, 362–382, https://doi.org/10.1016/j.jhydrol.2015.03.036, 2015.

Fisher, R. A.: Statistical Methods for Research Workers, in: Breakthroughs in Statistics: Methodology and Distribution, edited by Kotz, S. and Johnson, N. L., pp. 66–70, Springer, New York, NY, ISBN 978-1-4612-4380-9, https://doi.org/10.1007/978-1-4612-4380-9_6, 1992.

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrology and Earth System Sciences, 10, 535–552, https://doi.org/10.5194/hess-10-535-2006, publisher: Copernicus GmbH, 2006.

Fowler, K., Peel, M., Western, A., and Zhang, L.: Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function, Water Resources Research, 54, 3392–3408, https://doi.org/10.1029/2017WR022466, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2017WR022466, 2018.

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S.: Deep learning rainfall–runoff predictions of extreme events, Hydrology and Earth System Sciences, 26, 3377–3392, https://doi.org/10.5194/hess-26-3377-2022, publisher: Copernicus GmbH, 2022.

Garcia, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, Hydrological Sciences Journal, 62, 1149–1166, https://doi.org/10.1080/02626667.2017.1308511, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2017.1308511, 2017.

Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., and Veloso-Aguila, D.: The Central Chile Mega Drought (2010–2018): A climate dynamics perspective, International Journal of Climatology, 40, 421–439, https://doi.org/10.1002/joc.6219, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6219, 2020.

Griffin, D. and Anchukaitis, K. J.: How unusual is the 2012–2014 California drought?, Geophysical Research Letters, 41, 9017–9023, https://doi.org/10.1002/2014GL062433, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL062433, 2014.

Gu, L., Yin, J., Slater, L. J., Chen, J., Do, H. X., Wang, H.-M., Chen, L., Jiang, Z., and Zhao, T.: Intensification of Global Hydrological Droughts Under Anthropogenic Climate Warming, Water Resources Research, 59, e2022WR032 997, https://doi.org/10.1029/2022WR032997, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR032997, 2023.

Guo, D., Westra, S., and Maier, H. R.: Impact of evapotranspiration process representation on runoff projections from conceptual rainfall-runoff models, Water Resources Research, 53, 435–454, https://doi.org/10.1002/2016WR019627, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019627, 2017.

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.

Götte, J. and Brunner, M. I.: Hydrological Drought-To-Flood Transitions Across Different Hydroclimates in the United States, Water Resources Research, 60, e2023WR036 504, https://doi.org/10.1029/2023WR036504, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023WR036504, 2024.

Günther, D., Marke, T., Essery, R., and Strasser, U.: Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance, Water Resources Research, 55, 2779–2800, https://doi.org/10.1029/2018WR023403, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR023403, 2019.

Hammond, J., Anderson, B., Simeone, C., Brunner, M., Muñoz-Castro, E., Archfield, S., Magee, E., and Armitage, R.: Hydrological Whiplash: Highlighting the Need for Better Understanding and Quantification of Sub-Seasonal Hydrological Extreme Transitions, Hydrological Processes, 39, e70 113, https://doi.org/10.1002/hyp.70113, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.70113, 2025.

Handwerger, A. L., Huang, M.-H., Fielding, E. J., Booth, A. M., and Bürgmann, R.: A shift from drought to extreme rainfall drives a stable landslide to catastrophic failure, Scientific Reports, 9, 1569, https://doi.org/10.1038/s41598-018-38300-0, number: 1 Publisher: Nature Publishing Group, 2019.

He, X. and Sheffield, J.: Lagged Compound Occurrence of Droughts and Pluvials Globally Over the Past Seven Decades, Geophysical Research Letters, 47, e2020GL087 924, https://doi.org/10.1029/2020GL087924, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL087924, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803, 2020.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, Journal of Hydrology, 566, 595–606, https://doi.org/10.1016/j.jhydrol.2018.09.052, 2018.

Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, Hydrology and Earth System Sciences, 21, 3953–3973, https://doi.org/10.5194/hess-21-3953-2017, publisher: Copernicus GmbH, 2017.

Hughes, D., Birkinshaw, S., and Parkin, G.: A method to include reservoir operations in catchment hydrological models using SHETRAN, Environmental Modelling & Software, 138, 104 980, https://doi.org/10.1016/j.envsoft.2021.104980, 2021.

Hughes, D. A.: Adjusting rainfall inputs to hydrological models in a data-scarce area of Southern Africa, Hydrological Sciences Journal, 69, 1186–1197, https://doi.org/10.1080/02626667.2024.2360971, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2024.2360971, 2024.

Hurlbert, M. and Gupta, J.: The adaptive capacity of institutions in Canada, Argentina, and Chile to droughts and floods, Regional Environmental Change, 17, 865–877, https://doi.org/10.1007/s10113-016-1078-0, 2017.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth System Science Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, publisher: Copernicus GmbH, 2023a.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: Catchment attributes and hydro-meteorological time series for large-sample studies across hydrologic Switzerland (CAMELS-CH), https://doi.org/10.5281/zenodo.7784633, 2023b.

Jahromi, M. N., Miralles, D., Koppa, A., Rains, D., Zand-Parsa, S., Mosaffa, H., and Jamshidi, S.: Ten Years of GLEAM: A Review of Scientific Advances and Applications, in: Computational Intelligence for Water and Environmental Sciences, edited by Bozorg-Haddad, O. and Zolghadr-Asli, B., pp. 525–540, Springer Nature, Singapore, ISBN 978-981-19251-9-1, https://doi.org/10.1007/978-981-19-2519-1_25, 2022.

Kaufmann, J. and Schering, A.: Analysis of Variance ANOVA, in: Wiley StatsRef: Statistics Reference Online, John Wiley & Sons, Ltd, ISBN 978-1-118-44511-2, https://doi.org/10.1002/9781118445112.stat06938, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat06938, 2014.

Kiem, A. S. and Austin, E. K.: Drought and the future of rural communities: Opportunities and challenges for climate change adaptation in regional Victoria, Australia, Global Environmental Change, 23, 1307–1316, https://doi.org/10.1016/j.gloenvcha.2013.06.003, 2013.

Klemeš, V., V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, https://doi.org/10.1080/02626668609491024, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424-425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Knoben, W. J. M.: Setting expectations for hydrologic model performance with an ensemble of simple benchmarks, Hydrological Processes, 38, e15 288, https://doi.org/10.1002/hyp.15288, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.15288, 2024.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., and Woods, R. A.: A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments, Water Resources Research, 56, e2019WR025 975, https://doi.org/10.1029/2019WR025975, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR025975, 2020.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, publisher: Copernicus GmbH, 2018.

Le Moine, N.: Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit ?, phdthesis, Doctorat Géosciences et Ressources Naturelles, Université Pierre et Marie Curie Paris VI, https://hal.inrae.fr/tel-02591478, pages: 348, 2008.

Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, Water Resources Research, 35, 233–241, https://doi.org/10.1029/1998WR900018, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1998WR900018, 1999.

Lehner, B., Döll, P., Alcamo, J., Henrichs, T., and Kaspar, F.: Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis, Climatic Change, 75, 273–299, https://doi.org/10.1007/s10584-006-6338-4, 2006.

Lema, F., Mendoza, P. A., Vásquez, N. A., Mizukami, N., Zambrano-Bigiarini, M., and Vargas, X.: Technical note: What does the Standardized Streamflow Index actually reflect? Insights and implications for hydrological drought analysis, Hydrology and Earth System Sciences, 29, 1981–2002, https://doi.org/10.5194/hess-29-1981-2025, publisher: Copernicus GmbH, 2025.

Li, H., Xu, C. Y., and Beldring, S.: How much can we gain with increasing model complexity with the same model concepts?, Journal of Hydrology, 527, 858–871, https://doi.org/10.1016/j.jhydrol.2015.05.044, 2015.

Macdonald, E., Merz, B., Nguyen, V. D., and Vorogushyn, S.: Heavy-tailed flood peak distributions: what is the effect of the spatial variability of rainfall and runoff generation?, Hydrology and Earth System Sciences, 29, 447–463, https://doi.org/10.5194/hess-29-447-2025, publisher: Copernicus GmbH, 2025.

Magnusson, J., Gustafsson, D., Hüsler, F., and Jonas, T.: Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods, Water Resources Research, 50, 7816–7835, https://doi.org/10.1002/2014WR015302, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR015302, 2014.

Marengo, J. A. and Espinoza, J. C.: Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts, International Journal of Climatology, 36, 1033–1050, https://doi.org/10.1002/joc.4420, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.4420, 2016.

Martin, E. R.: Future Projections of Global Pluvial and Drought Event Characteristics, Geophysical Research Letters, 45, 11,913–11,920, https://doi.org/10.1029/2018GL079807, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL079807, 2018.

Matanó, A., de Ruiter, M. C., Koehler, J., Ward, P. J., and Van Loon, A. F.: Caught Between Extremes: Understanding Human-Water Interactions During Drought-To-Flood Events in the Horn of Africa, Earth's Future, 10, e2022EF002 747, https://doi.org/10.1029/2022EF002747, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022EF002747, 2022.

Matanó, A., Berghuijs, W. R., Mazzoleni, M., Ruiter, M. C. d., Ward, P. J., and Loon, A. F. V.: Compound and consecutive drought-flood events at a global scale, Environmental Research Letters, 19, 064 048, https://doi.org/10.1088/1748-9326/ad4b46, publisher: IOP Publishing, 2024.

Mathevet, T., Gupta, H., Perrin, C., Andréassian, V., and Le Moine, N.: Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds, Journal of Hydrology, 585, 124 698, https://doi.org/10.1016/j.jhydrol.2020.124698, 2020.

McClymont, K., Morrison, D., Beevers, L., and Carmen, E.: Flood resilience: a systematic review, Journal of Environmental Planning and Management, 63, 1151–1176, https://doi.org/10.1080/09640568.2019.1641474, publisher: Routledge _eprint: https://doi.org/10.1080/09640568.2019.1641474, 2020.

McMartin, D. W., Hernani Merino, B. H., Bonsal, B., Hurlbert, M., Villalba, R., Ocampo, O. L., Upegui, J. J. V., Poveda, G., and Sauchyn, D. J.: Limitations of Water Resources Infrastructure for Reducing Community Vulnerabilities to Extremes and Uncertainty of Flood and Drought, Environmental Management, 62, 1038–1047, https://doi.org/10.1007/s00267-018-1104-8, 2018.

McMillan, H. K.: Linking hydrologic signatures to hydrologic processes: A review, Hydrological Processes, 34, 1393–1409, https://doi.org/10.1002/hyp.13632, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.13632, 2020.

Melsen, L. A. and Guse, B.: Hydrological Drought Simulations: How Climate and Model Structure Control Parameter Sensitivity, Water Resources Research, 55, 10 527–10 547, https://doi.org/10.1029/2019WR025230, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR025230, 2019.

Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., Zappa, M., Mizukami, N., Mendoza, P. A., Clark, M. P., and Uijlenhoet, R.: Subjective modeling decisions can significantly impact the simulation of flood and drought events, Journal of Hydrology, 568, 1093–1104, https://doi.org/10.1016/j.jhydrol.2018.11.046, 2019.

Melsen, L. A., Puy, A., Torfs, P. J. J. F., and Saltelli, A.: The rise of the Nash-Sutcliffe efficiency in hydrology, Hydrological Sciences Journal, 70, 1248–1259, https://doi.org/10.1080/02626667.2025.2475105, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2025.2475105, 2025.

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., and Rajagopalan, B.: How do hydrologic modeling decisions affect the portrayal of climate change impacts?, Hydrological Processes, 30, 1071–1095, https://doi.org/10.1002/hyp.10684, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10684, 2016.

Merz, R., Miniussi, A., Basso, S., Petersen, K.-J., and Tarasova, L.: More Complex is Not Necessarily Better in Large-Scale Hydrological Modeling: A Model Complexity Experiment across the Contiguous United States, https://doi.org/10.1175/BAMS-D-21-0284.1, section: Bulletin of the American Meteorological Society, 2022.

MeteoSwiss: Federal Office for Meteorology and Climatology, https://www.meteoswiss.admin.ch/#tab=forecast-map, 2023.

Meylan, P., Favre, A.-C., and Musy, A.: Predictive Hydrology: A Frequency Analysis Approach, CRC Press, ISBN 978-1-57808-747-1, google-Books-ID: mpPa1ItTP2IC, 2012.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. a. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, publisher: Copernicus GmbH, 2011.

Mizukami, N., Clark, M. P., Gutmann, E. D., Mendoza, P. A., Newman, A. J., Nijssen, B., Livneh, B., Hay, L. E., Arnold, J. R., and Brekke, L. D.: Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled Forcing Data and Hydrologic Models, https://doi.org/10.1175/JHM-D-14-0187.1, section: Journal of Hydrometeorology, 2016.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., and Kumar, R.: On the choice of calibration metrics for "high-flow" estimation using hydrologic models, Hydrology and Earth System Sciences, 23, 2601–2614, https://doi.org/10.5194/hess-23-2601-2019, publisher: Copernicus GmbH, 2019.

Motavita, D., Chow, R., Guthke, A., and Nowak, W.: The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability, Journal of Hydrology, 573, 501–515, https://doi.org/10.1016/j.jhydrol.2019.03.054, 2019.

Muñoz-Castro, E., Mendoza, P. A., Vásquez, N., and Vargas, X.: Exploring parameter (dis)agreement due to calibration metric selection in conceptual rainfall–runoff models, Hydrological Sciences Journal, 68, 1754–1768, https://doi.org/10.1080/02626667.2023.2231434, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2023.2231434, 2023.

Muñoz-Castro, E., Anderson, B., and Brunner, M. I.: Implementation of four conceptual rainfall-runoff models to simulate drought-to-flood transitions in Chile and Switzerland, https://doi.org/10.5281/zenodo.17086604, 2025.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nasonova, O. N., Gusev, Y. M., and Kovalev, Y. E.: Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components, Hydrological Processes, 25, 1074–1090, https://doi.org/10.1002/hyp.7651, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.7651, 2011.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, publisher: Copernicus GmbH, 2015.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147–159, https://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, Hydrological Processes, 21, 435–446, https://doi.org/10.1002/hyp.6253, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.6253, 2007.

Peleg, N., Skinner, C., Fatichi, S., and Molnar, P.: Temperature effects on the spatial structure of heavy rainfall modify catchment hydromorphological response, Earth Surface Dynamics, 8, 17–36, https://doi.org/10.5194/esurf-8-17-2020, publisher: Copernicus GmbH, 2020.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Pizarro, A. and Jorquera, J.: Advancing objective functions in hydrological modelling: Integrating knowable moments for improved simulation accuracy, Journal of Hydrology, 634, 131 071, https://doi.org/10.1016/j.jhydrol.2024.131071, 2024.

Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V., and Perrin, C.: Process-based interpretation of conceptual hydrological model performance using a multinational catchment set, Water Resources Research, 53, 7247–7268, https://doi.org/10.1002/2016WR019991, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019991, 2017.

Pool, S., Vis, M., and Seibert, J.: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, Hydrological Sciences Journal, 63, 1941–1953, https://doi.org/10.1080/02626667.2018.1552002, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2018.1552002, 2018.

Probst, E. and Mauser, W.: Evaluation of ERA5 and WFDE5 forcing data for hydrological modelling and the impact of bias correction with regional climatologies: A case study in the Danube River Basin, Journal of Hydrology: Regional Studies, 40, 101 023, https://doi.org/10.1016/j.ejrh.2022.101023, 2022.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, Journal of Hydrology, 411, 66–76, https://doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420-421, 171–182, https://doi.org/10.1016/j.jhydrol.2011.11.055, 2012.

Quesada-Montano, B., Di Baldassarre, G., Rangecroft, S., and Van Loon, A. F.: Hydrological change: Towards a consistent approach to assess changes on both floods and droughts, Advances in Water Resources, 111, 31–35, https://doi.org/10.1016/j.advwatres.2017.10.038, 2018.

Rashid, M. M. and Wahl, T.: Hydrologic risk from consecutive dry and wet extremes at the global scale, Environmental Research Communications, 4, 071 001, https://doi.org/10.1088/2515-7620/ac77de, publisher: IOP Publishing, 2022.

Robeson, S. M.: Revisiting the recent California drought as an extreme value, Geophysical Research Letters, 42, 6771–6779, https://doi.org/10.1002/2015GL064593, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015GL064593, 2015.

Rouholahnejad Freund, E., Zappa, M., and Kirchner, J. W.: Averaging over spatiotemporal heterogeneity substantially biases evapotranspiration rates in a mechanistic large-scale land evaporation model, Hydrology and Earth System Sciences, 24, 5015–5025, https://doi.org/10.5194/hess-24-5015-2020, publisher: Copernicus GmbH, 2020.

Saavedra, D., Mendoza, P. A., Addor, N., Llauca, H., and Vargas, X.: A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments, Hydrological Processes, 36, e14 446, https://doi.org/10.1002/hyp.14446, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.14446, 2022.

Santos, L., Andréassian, V., Sonnenborg, T. O., Lindström, G., de Lavenne, A., Perrin, C., Collet, L., and Thirel, G.: Lack of robustness of hydrological models: a large-sample diagnosis and an attempt to identify hydrological and climatic drivers, Hydrology and Earth System Sciences, 29, 683–700, https://doi.org/10.5194/hess-29-683-2025, publisher: Copernicus GmbH, 2025.

Seibert, J. and Bergström, S.: A retrospective on hydrological catchment modelling based on half a century with the HBV model, Hydrology and Earth System Sciences, 26, 1371–1388, https://doi.org/10.5194/hess-26-1371-2022, publisher: Copernicus GmbH, 2022.

Seibert, J., Vis, M. J. P., Lewis, E., and van Meerveld, H.: Upper and lower benchmarks in hydrological modelling, Hydrological Processes, 32, 1120–1125, https://doi.org/10.1002/hyp.11476, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.11476, 2018.

Seibert, J., Staudinger, M., and van Meerveld, H. J. I.: Validation and Over-Parameterization—Experiences from Hydrological Modeling, in: Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives, edited by Beisbart, C. and Saam, N. J., pp. 811–834, Springer International Publishing, Cham, ISBN 978-3-319-70766-2, https://doi.org/10.1007/978-3-319-70766-2_33, 2019.

Sivakumar, B.: Global climate change and its impacts on water resources planning and management: assessment and challenges, Stochastic Environmental Research and Risk Assessment, 25, 583–600, https://doi.org/10.1007/s00477-010-0423-y, 2011.

Spieler, D. and Schütze, N.: Investigating the Model Hypothesis Space: Benchmarking Automatic Model Structure Identification With a Large Model Ensemble, Water Resources Research, 60, e2023WR036 199, https://doi.org/10.1029/2023WR036199, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023WR036199, 2024.

Stisen, S., Højberg, A. L., Troldborg, L., Refsgaard, J. C., Christensen, B. S. B., Olsen, M., and Henriksen, H. J.: On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes, Hydrology and Earth System Sciences, 16, 4157–4176, https://doi.org/10.5194/hess-16-4157-2012, publisher: Copernicus GmbH, 2012.

Sun, Q., Miao, C., AghaKouchak, A., and Duan, Q.: Century-scale causal relationships between global dry/wet conditions and the state of the Pacific and Atlantic Oceans, Geophysical Research Letters, 43, 6528–6537, https://doi.org/10.1002/2016GL069628, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL069628, 2016.

Swain, D. L., Langenbrunner, B., Neelin, J. D., and Hall, A.: Increasing precipitation volatility in twenty-first-century California, Nature Climate Change, 8, 427–433, https://doi.org/10.1038/s41558-018-0140-y, number: 5 Publisher: Nature Publishing Group, 2018.

Swain, D. L., Prein, A. F., Abatzoglou, J. T., Albano, C. M., Brunner, M., Diffenbaugh, N. S., Singh, D., Skinner, C. B., and Touma, D.: Hydroclimate volatility on a warming Earth, Nature Reviews Earth & Environment, 6, 35–50, https://doi.org/10.1038/s43017-024-00624-z, publisher: Nature Publishing Group, 2025.

Tabari, H., Hosseinzadehtalaei, P., Thiery, W., and Willems, P.: Amplified Drought and Flood Risk Under Future Socioeconomic and Climatic Change, Earth's Future, 9, e2021EF002 295, https://doi.org/10.1029/2021EF002295, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002295, 2021.

Tallaksen, L. M., Madsen, H., and Clausen, B.: On the definition and modelling of streamflow drought duration and deficit volume, Hydrological Sciences Journal, 42, 15–33, https://doi.org/10.1080/02626669709492003, publisher: Taylor & Francis, 1997.

Tang, G., Clark, M. P., and Papalexiou, S. M.: SC-Earth: A Station-Based Serially Complete Earth Dataset from 1950 to 2019, Journal of Climate, 34, 6493–6511, https://doi.org/10.1175/JCLI-D-21-0067.1, publisher: American Meteorological Society Section: Journal of Climate, 2021.

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., Beck, H. E., Wood, A. W., Newman, A. J., and Papalexiou, S. M.: The Impact of Meteorological Forcing Uncertainty on Hydrological Modeling: A Global Analysis of Cryosphere Basins, Water Resources Research, 59, e2022WR033 767, https://doi.org/10.1029/2022WR033767, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR033767, 2023.

Tang, G., Clark, M. P., Knoben, W. J. M., Liu, H., Gharari, S., Arnal, L., Wood, A. W., Newman, A. J., Freer, J., and Papalexiou, S. M.: Uncertainty Hotspots in Global Hydrologic Modeling: The Impact of Precipitation and Temperature Forcings, Bulletin of the American Meteorological Society, https://doi.org/10.1175/BAMS-D-24-0007.1, section: Bulletin of the American Meteorological Society, 2025.

Thirel, G., Santos, L., Delaigue, O., and Perrin, C.: On the use of streamflow transformations for hydrological model calibration, Hydrology and Earth System Sciences, 28, 4837–4860, https://doi.org/10.5194/hess-28-4837-2024, publisher: Copernicus GmbH, 2024.

Tuel, A., Schaefli, B., Zscheischler, J., and Martius, O.: On the links between sub-seasonal clustering of extreme precipitation and high discharge in Switzerland and Europe, Hydrology and Earth System Sciences, 26, 2649–2669, https://doi.org/10.5194/hess-26-2649-2022, publisher: Copernicus GmbH, 2022.

Valenzuela, R., Garreaud, R., Vergara, I., Campos, D., Viale, M., and Rondanelli, R.: An extraordinary dry season precipitation event in the subtropical Andes: Drivers, impacts and predictability, Weather and Climate Extremes, 37, 100 472, https://doi.org/10.1016/j.wace.2022.100472, 2022.

Valéry, A., Andréassian, V., and Perrin, C.: Regionalization of precipitation and air temperature over high-altitude catchments – learning from outliers, Hydrological Sciences Journal, 55, 928–940, https://doi.org/10.1080/02626667.2010.504676, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2010.504676, 2010.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments, Journal of Hydrology, 517, 1166–1175, https://doi.org/10.1016/j.jhydrol.2014.04.059, 2014a.

Valéry, A., Andréassian, V., and Perrin, C.: 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments, Journal of Hydrology, 517, 1176–1187, https://doi.org/10.1016/j.jhydrol.2014.04.058, 2014b.

van Kempen, G., van der Wiel, K., and Melsen, L. A.: The impact of hydrological model structure on the simulation of extreme runoff events, Natural Hazards and Earth System Sciences, 21, 961–976, https://doi.org/10.5194/nhess-21-961-2021, publisher: Copernicus GmbH, 2021.

Van Loon, A. F. and Van Lanen, H. a. J.: A process-based typology of hydrological drought, Hydrology and Earth System Sciences, 16, 1915–1946, https://doi.org/10.5194/hess-16-1915-2012, publisher: Copernicus GmbH, 2012.

Visconti, G.: After the Flood: Disasters, Ideological Voting and Electoral Choices in Chile, Political Behavior, 44, 1985–2004, https://doi.org/10.1007/s11109-022-09814-1, 2022.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resources Research, 41, https://doi.org/10.1029/2004WR003059, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004WR003059, 2005.

Vásquez, N., Cepeda, J., Gómez, T., Mendoza, P. A., Lagos, M., Boisier, J. P., Álvarez Garretón, C., and Vargas, X.: Catchment-Scale Natural Water Balance in Chile, in: Water Resources of Chile, edited by Fernández, B. and Gironás, J., pp. 189–208, Springer International Publishing, Cham, ISBN 978-3-030-56901-3, https://doi.org/10.1007/978-3-030-56901-3_9, 2021.

Wang, S.-Y. S., Yoon, J.-H., Becker, E., and Gillies, R.: California from drought to deluge, Nature Climate Change, 7, 465–468, https://doi.org/10.1038/nclimate3330, number: 7 Publisher: Nature Publishing Group, 2017.

Ward, P. J., de Ruiter, M. C., Mård, J., Schröter, K., Van Loon, A., Veldkamp, T., von Uexkull, N., Wanders, N., AghaKouchak, A., Arnbjerg-Nielsen, K., Capewell, L., Carmen Llasat, M., Day, R., Dewals, B., Di Baldassarre, G., Huning, L. S., Kreibich, H., Mazzoleni, M., Savelli, E., Teutschbein, C., van den Berg, H., van der Heijden, A., Vincken, J. M. R., Waterloo, M. J., and Wens, M.: The need to integrate flood and drought disaster risk reduction strategies, Water Security, 11, 100 070, https://doi.org/10.1016/j.wasec.2020.100070, 2020.

Wasti, A., Ray, P., Wi, S., Folch, C., Ubierna, M., and Karki, P.: Climate change and the hydropower sector: A global review, WIREs Climate Change, 13, e757, https://doi.org/10.1002/wcc.757, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.757, 2022.

Wilcoxon, F.: Individual Comparisons by Ranking Methods, Biometrics Bulletin, 1, 80–83, https://doi.org/10.2307/3001968, publisher: [International Biometric Society, Wiley], 1945.

Wu, S., Zhao, J., and Sivapalan, M.: A parsimonious daily water balance model based on the proportionality hypothesis, Journal of Hydrology, 649, 132 414, https://doi.org/10.1016/j.jhydrol.2024.132414, 2025.

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

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, https://doi.org/10.1029/2007WR006716, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR006716, 2008.

Zhao, F., Nie, N., Liu, Y., Yi, C., Guillaumot, L., Wada, Y., Burek, P., Smilovic, M., Frieler, K., Buechner, M., Schewe, J., and Gosling, S. N.: Benefits of Calibrating a Global Hydrological Model for Regional Analyses of Flood and Drought Projections: A Case Study of the Yangtze River Basin, Water Resources Research, 61, e2024WR037 153, https://doi.org/10.1029/2024WR037153, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2024WR037153, 2025.