Analyzing the informative value of alternative hazard indicators for monitoring drought risk for human water supply and river ecosystems at the global scale

Claudia Herbert[1] and Petra Döll[1,2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, 60438, Germany
[2]Senckenberg Leibniz Biodiversity and Climate Research Centre Frankfurt (SBiK-F), Frankfurt am Main, 60325, Germany

Correspondence to: Claudia Herbert (c.herbert@em.uni-frankfurt.de)

Abstract

Streamflow drought hazard indicators (SDHIs) are mostly lacking in large-scale drought early warning systems (DEWS). This paper presents a new systematic approach for selecting and computing SDHIs for monitoring drought risk for human water supply from surface water and for river ecosystems that is also relevant for meteorological or soil moisture drought. We recommend considering the habituation of ~~people and ecosystems~~the system at risk to the streamflow regime (e.g., to a certain interannual variability or a certain relative reduction of streamflow) when selecting indicators. Distinguishing four indicator types, we classify indicators of drought magnitude (~~e.g.,~~ water anomaly during a pre-defined period) and severity (cumulated magnitude since the onset of the drought event) and specify the many relevant decisions that need to be made when computing SDHIs and other drought hazard indicators. Using the global hydrological model WaterGAP2.2d, w~~W~~e quantify ~~nine~~ eight existing and three new SDHIs globally ~~using the global hydrological model WaterGAP2.2d~~. For large-scale DEWS based on the output of hydrological models, we recommend ~~selected~~ specific SDHIs that are suitable for assessing the drought risk of 1) river ecosystems, 2) water users without access to large reservoirs, 3) water users with access to large reservoirs, and 4) that are suitable for informing reservoir managers. ~~that are differently adapted to low water availability, characterized by either perennial or intermittent streamflow regime, and with or without access to large reservoirs.~~ These SDHIs include both drought magnitude and severity indicators that differ by the temporal averaging period and the habituation of the risk system to reduced water availability. Depending on the habituation of the risk system, d~~D~~rought magnitude is best quantified either by the relative deviation from the mean or by the return period of the streamflow value that is based on the frequency of non-exceedance. To compute the return period, w~~based on empirical percentiles or and relative deviation from mean. W~~e favor empirical percentiles over the standardized streamflow indicator as the former do not entail uncertainties due to the fitting of a probability distribution and can be computed for all streamflow time series. Drought severity should be assessed with indicators that imply habituation to a certain degree of interannual variability, to a certain reduction from mean ~~monthly~~ streamflow, and to the ability to fulfill ~~seasonally varying~~ human water demand ~~from surface water~~ and environmental flows. ~~, and severity by return period or water volume below a threshold relative to mean annual streamflow. Both anomaly and~~

deficit indicators should be provided. Reservoir managers are best informed by the SDHIs of the grid cell that represents inflow into the reservoir. The DEWS must provide comprehensive and clear explanations about the suitability of the provided indicators for specific risk systems.

1 Introduction

Drought occurs when there is a prolonged time period with less water than normal in different components of the hydrological cycle (van Loon et al., 2016) but the term drought also has the connotation that during the drought period there is less water than required (Popat and Döll, 2021). No universal definition of "drought" exists (Lloyd-Hughes, 2014). While drought is a local to regional phenomenon, its impacts can have transnational to global dimensions, in particular related to crop production and trade (Wilhite and Glantz, 1985; van Loon, 2015; UNECE, 2015). Streamflow drought in transboundary basins implies direct international impacts. Hence, global-scale assessment, monitoring, and forecasting of drought hazards or risks have the potential to support drought risk management (Pozzi et al., 2013).

Drought poses numerous risks to humans and ecosystems. A specific drought risk is a function of hazard, exposure, and vulnerability, while the term "drought impact" relates to the manifested risk (Field and Barros, 2014). In general, the term "hazard" refers to the physical event, "exposure" to the presence of people or ecosystems that could be negatively affected, and "vulnerability" to the susceptibility of a system to drought impacts and its (short term) coping and (long term) adaptation capacity (Field and Barros, 2014).

Drought risk indicators, and thus drought hazard, exposure and vulnerability indicators, should be designed specifically for the targeted risk (Lloyd-Hughes, 2014; Spinoni et al., 2018; Hagenlocher et al., 2019). As an example, the risk of "not being able to provide enough water to fulfill the customers' water demand during times of lower water availability than normal" constitutes a drought risk for water supply companies. If the water supply source is a river, a suitable drought hazard indicator should be based on streamflow data or on water storage in the upstream reservoir. However, a very large number of drought hazard indicators has been proposed and applied by experts without stringent consideration of the targeted risks (Wilhite and Glantz, 1985; Lloyd-Hughes, 2014). To identify a suitable risk-specific hazard indicator, the first step is to decide which water flow or storage should be taken into account, e.g. precipitation, soil moisture or streamflow, as the determined drought hazards depend on the considered physical variable (Satoh et al., 2021). Even after determining the appropriate risk-specific physical variable, e.g. streamflow, there is still a large choice of possible drought hazard indicators to quantify the occurrence or severity of streamflow droughts (WMO and GWP, 2016; Yihdego et al., 2019). A stakeholder survey encompassing 33 regional to global drought early warning systems (DEWS) revealed that streamflow drought hazard indicators (SDHIs) are rarely applied in DEWS, while drought hazard indicators based on meteorological variables, soil moisture, and

remotely-sensed vegetation conditions dominate. Among SDHIs, streamflow percentiles are mostly applied, e.g. in the US Drought Monitor. Other indicators include the Palmer Hydrological Drought Severity Index (Palmer, 1965), cumulative streamflow anomalies (Fleig et al., 2006; Lehner et al., 2006; van Loon et al., 2012; Heudorfer and Stahl, 2017), and the standardized streamflow (Modarres, 2007; Nalbantis and Tsakiris, 2009) or runoff index (Shukla and Wood, 2008; Satoh et al., 2021). At the continental scale, only the European Drought Monitor provides an SDHI (Cammalleri et al., 2016a), which has also been tested for global implementation in the Global Drought Observatory (Cammalleri et al., 2020). There is currently no global-scale operational streamflow drought hazard monitoring system.

~~Previous research has revealed that there is often no common understanding among stakeholders about drought hazard concepts~~ (Steinemann et al., 2015)~~. Also, in most descriptions of drought indicator calculations it is not made explicit what is assumed to be "normal". For instance, defining the long-term mean value of the physical variable per calendar month as normal state implies that people and ecosystems are habituated to the seasonality of water availability. Applying percentiles per calendar month instead implies the habituation to interannual variability. Clearly, the conception or selection of hazard indicators needs to take into account the habituation and thus vulnerability of the system at risk. However, investigations and guidance on how to select the optimal SDHI, considering both the targeted risk and the habituation of the system at risk to the streamflow regime, are missing.~~

~~Streamflow drought hazard can be estimated using either observed or modeled streamflow data. If no such data are available, streamflow drought hazard is estimated by applying meteorological indicators such as the standardized precipitation index SPI (McKee et al., 1993) and the standardized precipitation potential evaporation index SPEI (Vicente-Serrano et al., 2010), where the delayed response of streamflow to below normal precipitation is considered through longer averaging periods, i.e., by comparing mean precipitation conditions of the preceding n months to the respective n months of the reference period (SPIn). Averaging periods can range between 1 and 24 months (Gevaert et al., 2018). However, studies have shown that meteorological indicators have limitations in describing hydrological drought processes and suggest including streamflow drought indicators in drought management (Haslinger et al., 2014; Blauhut et al., 2016; Laaha et al., 2017). Where streamflow observations are not available, hydrological models can compute the response of streamflow to precipitation and other climatic variables to determine spatially and temporally continuous drought hazard indicators that take into account the different characteristics of the river basins (Lehner et al., 2006). Still, the meteorological variables precipitation and potential evapotranspiration are known to be major drivers of streamflow and its variability, and hydrological models, in particular large-scale models, suffer from significant uncertainties such that the added value of simulated SDHIs should be assessed.~~

SDHIs are commonly classified into threshold-based and standardized indicators (van Loon, 2015). The threshold level method (TLM) was first applied by Yevjevich (1967), who defined that a drought event begins when streamflow falls below a certain threshold (e.g. a percentile) and ends as soon as the threshold is exceeded. Then, drought magnitude is the streamflow deficit at the considered ~~time~~ period (computed as the difference between the threshold streamflow and the actual streamflow in that ~~time~~ period), while drought severity is equivalent to the cumulative magnitude since the beginning of the drought event. Standardized indicators such as the standardized precipitation index (SPI) (McKee et al., 1993) ~~and t~~The standardized

streamflow ~~indicator~~ index (SSI) (Zaidman et al., 2002; Modarres, 2007; Nalbantis and Tsakiris, 2009) quantifyies the anomaly of the variable (e.g. precipitation or streamflow) during a certain ~~time~~ period from the long-term mean ~~streamflow~~ in units of standard deviation ~~and is computed like the SPI~~. Negative values quantify the drought magnitude per time step. ~~SSI has been applied using a 1-month averaging period (SSI1) (Zaidman et al., 2002; Modarres, 2007; Nalbantis and Tsakiris, 2009) as well as longer averaging periods (SSI3, SSI6, SSI12) (Svensson et al., 2017; Wan et al., 2021).~~ However, classification in threshold-based and standardized indicators is somewhat misleading, since standardized indicators can also be cumulated to derive drought severity, which requires setting of a threshold as is the case for TLM indicators (McKee et al., 1993; Barker et al., 2019, van Oel et al., 2018, Tijdeman et al., 2020). On the other hand, comparing SSI and threshold-based indicators directly implies that different drought characteristics (magnitude and severity) are analysed. Moreover, the term drought severity is sometimes used to describe drought magnitude and vice versa (Steinemann et al., 2015, Vidal et al., 2009; López-Moreno et al., 2009). Certainly, an improved classification of drought hazard indicators would facilitate a better understanding of drought characteristics and ~~provide guidance~~guide ~~in selecting~~the selection of appropriate drought hazard indicators.

~~A further consideration in designing SDHIs is how to conceptualize drought in intermittent or highly seasonal streamflow regimes. If periods of zero flow are a normal part of the streamflow regime, as it is the case in arid regions, then it is meaningless to assess streamflow deficits during these periods. Hence, arid regions are often excluded from global drought analyses (Corzo Perez et al., 2011; Prudhomme et al., 2014; Spinoni et al., 2019). Some authors tested rather high percentiles as thresholds to characterize drought in intermittent streamflow regimes (e.g. the 80th percentile or Q20, the streamflow that is exceeded in two out of ten months) (Woo and Tarhule, 1994; Tate and Freeman, 2000; Fleig et al., 2006). This approach, however, has been criticized as it is not consistent with the anomaly concept of drought (van Huijgevoort et al., 2012). To overcome these limitations, van Huijgevoort et al. (2012) introduced a method to identify streamflow drought at the global scale that is also applicable for intermittent rivers. It that combines the TLM with the consecutive dry period method (CDPM) for streamflow, in analogy to the consecutive dry days (CDD) indicator for precipitation (Vincent and Mekis, 2006; Griffiths and Bradley, 2007). Using this combined method, a drought in a period with streamflow identified with the TLM is allowed to continue in a subsequent zero flow period. Very short periods of zero flow are excluded from the assessment. Although more sophisticated compared to the TLM alone, the combined method as described in van Huijgevoort et al. (2012) which, however, may be too complex to be applied in DEWS. Moreover, the final scaling procedure of percentiles in months where both TLM and CDPM apply might result in thresholds that are not intuitive.~~

Previous research has revealed that there is often no common understanding among stakeholders about drought hazard concepts (Steinemann et al., 2015). Also, in most descriptions of drought indicator calculations, it is not made explicit what is assumed to be "normal". For instance, defining the long-term mean value of the physical variable per calendar month as normal state implies that people and ecosystems are habituated and thus adapted to the seasonality of water availability. Applying percentiles per calendar month instead implies the habituation to interannual variability. Clearly, the conception or selection of hazard indicators needs to take into account the habituation and ~~thus~~the related vulnerability of the system at risk.

However, investigations and guidance on how to select the optimal SDHI, considering both the targeted risk and the habituation of the system at risk to the streamflow regime, are missing.

A further consideration in designing SDHIs is how to conceptualize drought in intermittent or highly seasonal streamflow regimes. If periods of zero flow are a normal part of the streamflow regime, as ~~it~~ is the case in arid regions, then it is meaningless to assess streamflow deficits during these periods. Hence, arid regions are often excluded from global drought analyses (Corzo Perez et al., 2011; Prudhomme et al., 2014; Spinoni et al., 2019). To overcome these limitations, van Huijgevoort et al. (2012) introduced a method that combines the TLM with the consecutive dry period method (CDPM) for streamflow, in analogy to the consecutive dry days (CDD) indicator for precipitation (Vincent and Mekis, 2006; Griffiths and Bradley, 2007). However, this method may be too complex to be applied in DEWS.

This paper analyzes which SDHIs are suitable for assessing and monitoring drought risk for human water supply from surface water and for river ecosystems in large-scale DEWS. We propose a systematic approach to indicator selection, which encompasses the explicit consideration of the habituation of people and river ecosystems to streamflow availability as well as a new classification system for drought hazard indicators. This new methodology is exemplified at the global scale for eight existing and three newly developed SDHIs using a) modeled output from ~~Applying~~ the global water resources and use model WaterGAP2.2d and b) observed monthly streamflow at four selected gauging stations.~~for the reference period 1986-2015, we compare drought hazard globally as determined by nine existing and three newly developed hazard indicators.~~

The following section describes how streamflow and other variables required for the computation of the SDHIs were computed and defines the ~~twelve~~ eleven investigated SDHIs. In Sect. 3, we present the new systematic approach for selecting and computing SDHIs. ~~and illustrate the approach using observed streamflow at two gauging stations.~~ In Sect. 4, we analyze spatial and temporal discrepancies and similarities of the indicators at the global scale. In Sect. 5, we give recommendations on the general suitability of the indicators as well as for large-scale applications. Finally, we draw conclusions in Sect. 6.

2 Methods and data

2.1 Streamflow data

2.1.1 Streamflow observations

Eight SDHIs were computed for four selected gauging stations using monthly streamflow data from the Global Runoff Data Centre (GRDC, 2019) for the period 1986-2015 (Figs. 6, S2, and S6). The stations comprise the Danube River at Hofkirchen (Germany), the Angara River at Boguchany (Russia), the White River near Oacoma (U.S.), and the Orange River at Vioolsdrif (South Africa). Moreover, a limited model validation was performed (~~S~~supplementary material S2) using monthly streamflow data from 220 GRDC stations with continuous time series during the reference period 1986-2015. The model

validation focused on the correlation between observed Q80 per calendar month (the streamflow that is exceeded in eight out of ten months) and Q80 as modeled by WaterGAP.*

2.1.2 Modeled streamflow

~~Twelve~~The~~Eleven~~ SDHIs ~~(Table 1)~~ were computed for the whole land area except Greenland and Antarctica with a spatial resolution of 0.5° using monthly time series of WaterGAP 2.2d model output for the reference period 1986-2015 (Sect. 2.2). For computing each indicator, we used the 30 monthly values available for each of the 12 calendar months individually to determine distributions, thresholds, and deficits. All indicators were computed using streamflow of the standard model run (Qant) ("ant": anthropogenic), in which the impact of human water use and man-made reservoirs on streamflow is simulated. Naturalized ("nat") streamflow (Qnat) without these two types of human activities was only used for deriving environmental streamflow requirements for the indicator CQDI1(WUs-EFR) (Sect. 2.3.5).

~~Moreover, a limited model validation was performed for WaterGAP 2.2d calibration stations using monthly streamflow observations provided by the Global Runoff Data Centre (GRDC, 2019) for the period 1986-2015 (Supplementary material S4).~~

2.~~1~~ 2 Global-scale simulation of streamflow and~~,~~ surface water use ~~and PET.~~

~~Hydrological drought hazard indicators~~SDHIs were computed using output from the global water availability and water use model WaterGAP2.2d (Müller Schmied et al., 2021). WaterGAP2.2d has a spatial resolution of 0.5 degrees latitude by 0.5 degrees longitude (55 km × 55 km at the equator) and covers the whole global land area except Antarctica. WaterGAP consists of the WaterGAP Global Hydrology Model (WGHM) and five water use models for the sectors households, manufacturing, and cooling of thermal power plants (Flörke et al., 2013) as well as irrigation and livestock. WGHM computes daily time series of fast surface and subsurface runoff, groundwater recharge, and streamflow as well as water storage variations in canopy, snow, soil, groundwater, lakes, reservoirs, wetlands, and rivers. Model input includes time series of climate data between 1901 and 2016 and physio-geographic information, such as land cover, soil type, relief, and hydrogeology. For this study, WaterGAP 2.2d was forced by the WFDEI-GPCC climate data set (Weedon et al., 2014), which was developed by applying the forcing data methodology from the EU project WATCH on ERA-Interim reanalysis data. ~~Potential evapotranspiration (PET), required for the calculation of SPEI12, was computed using the Priestley Taylor equation. In addition to the standard model run ("ant": anthropogenic), in which the impact of human water use and man-made reservoirs on streamflow is simulated, naturalized ("nat") conditions were computed by turning off these two types of human activities.~~ Daily model outputs of ~~anthropogenic and naturalized~~ streamflow ~~(Qant and Qnat), PET,~~ and surface water abstractions (WUs) were aggregated to monthly time series. WaterGAP total runoff is calibrated against long-term mean annual streamflow at 1319 gauging stations worldwide covering approximately 54% of the Earth's land area (except Greenland and Antarctica). A detailed model description and evaluation can be found in Müller Schmied et al. (2021). Please note that while WaterGAP simulates the impact

of reservoirs on streamflow, the accuracy is very low as it is unknown how all man-made reservoirs on Earth are managed such that a generic algorithm is used to simulate human reservoir management decisions.

In several model intercomparison and assessment studies, WaterGAP proved suitable for computing streamflow and SDHIs, although the discrepancies between simulated and observed low flows, seasonality, and interannual variability can be significant at the regional scale (see literature review in the supplementary material S1). A limited model validation of the WaterGAP version 2.2d applied in this study (supplementary material S~~3~~2) revealed ~~The analysis reveals~~ that Q80 is overestimated by WaterGAP in 63% of all months and stations ~~and in 53% if only relevant deviations > 10% are considered. The~~with median percent deviations ~~ranges~~ between 35% in February and -7% in July (Fig. S1). In another model validation exercise, SSI3 as modeled by WaterGAP was compared to observed SSI3 at 183 gauging stations (supplementary material S2). With a median NSE of 0.5 and an interquartile range of 0.2-0.7, WaterGAP 2.2d model output showed a moderate agreement with the observations. NSE exceeded 0.7 at 25 out of the 183 stations mainly located in central and eastern Europe and the United States. ~~Figure 3 indicates a tendency of WaterGAP 2.2d to overestimate observed Q80 between October and April, while Q80 during the low-flow period in the Northern Hemisphere (May to September) is better captured.~~

~~In several model intercomparison studies, WaterGAP was often among the best performing global hydrological models (GHMs). Kumar et al. (2022) assessed the ability of nine catchment scale models and eight GHMs to simulate hydrological droughts in eight large catchments around the world. Comparing simulated and observed streamflow deficits and SSI1 (SRI) (their Tables 2 and 3), WaterGAP is among the two to three best performing GHMs with performance indicators (R^2 and Nash Sutcliffe efficiency) comparable to those of the catchment scale models. In another assessment of streamflow drought based on observed or simulated streamflow at 293 locations in Europe (Tallaksen and Stahl, 2014), WaterGAP performed well as compared to six other GHMs. In their Fig. 3, WaterGAP results are better than the multi model median for all four performance measures. Moreover, WaterGAP performed best regarding the simulation of drought persistence (their Fig. 4). Prudhomme et al. (2011) analyzed the ability of three GHMs to reproduce historical streamflow drought events in European basins using the regional deficiency index (RDI). While all three models are found to broadly capture the spatiotemporal drought development, the authors conclude that WaterGAP "is arguably best suited to reproduce most regional characteristics of large scale high and low flow events in Europe" (Prudhomme et al., 2011: 1181). However, WaterGAP tends to overestimate the variability in RDI, which is explained by insufficient soil storage capacity. In an intercomparison study among six GHMs (Zaherpour et al., 2018), WaterGAP showed the best results in simulating monthly streamflow in 27 out of 40 river basins worldwide and in each of the eight hydrobelts (their Fig. 2 and Table 3). In five out of eight hydrobelts, the mean weighted absolute error of Q95 was lowest for WaterGAP. Nevertheless, the study revealed that WaterGAP tends to overestimate low flows, and that discrepancies between simulated and observed seasonality and interannual variability can be significant. In a different multi-model validation study based on five global hydrological and land surface models (Veldkamp et al., 2018), WaterGAP was the only model that slightly underestimated variability in monthly streamflow while the others overestimated variability. Correlation with observed monthly streamflow though was highest for WaterGAP in both managed and near natural basins across the globe (their Fig. 3h). Döll et al. (2016) compared monthly low flow Q90 as computed by the~~

GHMs WaterGAP and PCRGLOB-WB to observations at 821 WaterGAP calibration stations across the globe. Overall, low flows could be simulated with reasonable accuracy by both GHMs and were overestimated at most stations. WaterGAP results showed a better fit to observations since it is calibrated against mean annual streamflow at the considered stations (their Fig. 3). Despite calibration, WaterGAP simulations show a lower fit to small observed Q90 values below 1 $km^3\ month^{-1}$.

2.2 3 SDHIs2.2.1 The meteorological indicators standardized precipitation index SPI and standardized precipitation-potential evaporation index SPEIStandardized meteorological indicators SPI and SPEI

SPI and SPEI time series were computed from gridded estimates of monthly precipitation and the difference of monthly precipitation and potential evapotranspiration from the WaterGAP model. SPI is typically computed by averaging the precipitation over n previous months, termed averaging periods, and fitting the averaged precipitation to a nonlinear parametric statistical distribution, separately for each calendar month. The probabilities of the precipitation values are transformed to a variable Z with a normal distribution that has a mean of zero and a standard deviation of 1 (McKee et al., 1993, Stagge et al., 2015), using an approximation method introduced by Abramowitz and Stegun (1965). The value of the variable Z (also called z score) is equal to the value of the SPI. Thus, an SPI of -1 describes a precipitation value that is one standard deviation lower than the mean precipitation of the calendar month. The mean of the normal distribution is equal to the median of the fitted non-linear cumulative distribution function (Vicente-Serrano et al., 2010).

SPI time series were computed at the global scale following the method described in . Monthly precipitation data are first fitted to a probability distribution (e.g. gamma or Pearson Type III) and then transformed to the standard normal random variable Z (Eq. 1) (also termed z score), which is the SPI, following an approximation method introduced by Abramowitz and Stegun (1965). The standard normal distribution is characterized by a mean of zero and standard deviation of 1. A value of -1, for example, indicates that the precipitation value deviates from the long-term mean by one standard deviation.

$$Z = (X -)/ \quad (1)$$

with X = variable (e.g. precipitation), = mean, and = standard deviation.

The SPI can be quantified for different averaging periods of typically 1 to 36 months. In the present study, SPI time series were computed using a 12-month averaging period (SPI12), which is recommended for the assessment of hydrological drought impacts (WMO and GWP, 2016). For a limited sensitivity analysis, SPI time series with averaging periods of 3, 6, 9, and 12 months were derived for 218 WaterGAP calibration stations (Sect. 4.2.5Supplementary material S4). The indicator was computed using the SCI package for R (Gudmundsson and Stagge, 2016), fitting a gamma distribution to the precipitation time series. SPEI12 time series were also computed using the SCI package for R, however, applying the Log-logistics distribution as recommended for the SPEI

SPEI12 time series were calculated at the global scale according to the method presented in Vicente-Serrano et al. (2010) using a 12-month averaging period. Similar to SPI, the SCI package for R was utilized, however, applying the Log-logistics distribution as recommended for the SPEI (Vicente-Serrano et al., 2010; Vicente-Serrano and Beguería, 2016).

2.23.2 1 Standardized streamflow anomaly indicators SSI1 and SSI12

SSI1 was computed using mean monthly streamflow for Qant analogously to SPI1 (McKee et al., 1993) following the method provided in Kumar et al. (2009) using mean monthly streamflow for each of the 12 calendar months. First, the 30 monthly streamflow values per calendar month were fitted to the gamma distribution using the R package fitdistrplus. The probabilities of the streamflow values were transformed to a variable Z with a normal distribution that has a mean of zero and a standard

deviation of 1 (McKee et al., 1993, Stagge et al., 2015), using an approximation method introduced by Abramowitz and Stegun (1965). The value of the variable Z (also called z-score) is equal to the value of the SSI1. Thus, an SSI1 of -1 describes a streamflow value that is one standard deviation lower than the mean streamflow of the calendar month. The mean of the normal distribution is equal to the median of the fitted non-linear cumulative distribution function (Vicente-Serrano et al., 2010). ~~We applied the R package fitdistrplus and t~~The gamma distribution ~~, which~~ showed the best fit among 23 parametric probability distributions for most grid cells. The goodness-of-fit between simulated streamflow values and the probability distribution was assessed based on the one-sample Kolmogorov–Smirnov test (KS test) at the 0.05 significance level. The fits were rejected in 17% to 21% of all grid cells (excluding Greenland) depending on the calendar month.

SSI12 ~~is~~ was computed like SSI1, but with an averaging period of 12 months. For SSI12, the fits were rejected in around 6% of all grid cells (excluding Greenland) with only slight variations among the calendar months.

2.~~2~~3.~~3 2~~ Cumulative streamflow deficit indicators CQDI1~~-~~(Q50), CQDI1~~-~~(Q80), ~~and~~ CQDI1~~-~~(Q80-HS), and CQDI6(Q80)

~~CQDI1-Q50~~CQDI1(Q50) is the cumulative, volume-based streamflow deficit computed following the threshold level method (TLM) (Sect. 1). It should be noted that the term "deficit", which is generally used for the TLM, refers to the negative anomaly below a selected threshold, and not to an unsatisfied water demand. With ~~CQDI1-Q50~~CQDI1(Q50), a deficit is defined to occur if modeled monthly streamflow is lower than the 50th percentile (median) of the long-term ~~mean~~ calendar month streamflow (Eq. 1). The empirical percentile Q50 was computed in R using the quantile function with the default quantile algorithm. The streamflow deficit is computed as~~:~~

$$\text{streamflow deficit}_{m,y} = Q50_m - Q_{m,y} \quad (\text{for } Q_{m,y} < Q50_m) \quad (1)$$

with m = month, y = year, ~~deficit$_{m,y}$ = current monthly streamflow deficit,~~ $Q50_m$ = calendar month median, and $Q_{m,y}$ = current streamflow.

~~In each month, the streamflow deficit volume is calculated as the difference between the median of all 30 calendar month streamflow values during the reference period and the water volume that was actually transported in the stream in this month.~~ The last deficit month is the last month of the drought event. Monthly deficits (drought magnitude) are accumulated for all drought months to obtain severity. ~~Any streamflow surplus over the median in a single month between two deficit months does not decrease the cumulative deficit value.~~ The cumulative streamflow deficit (in units of m^3) is normalized by mean annual streamflow (in units of m^3). A value of 2 [-], for example, indicates that the cumulative streamflow deficit in a certain month is twice the mean annual streamflow. Following Spinoni et al. (2019), a drought event is defined to start with at least two consecutive months with a deficit and it ends (deficit set to zero) if there are two consecutive months without a defi-

cit (two months criterion, 2mc). This approach avoids that short-term streamflow deficits that hardly pose a drought hazard to humans and other biota are defined as drought events (Spinoni et al., 2019). Any streamflow surplus over the median in a single month between two deficit months does not decrease the cumulative deficit value. Q50 as a rather high threshold can be viewed as a "conservative upper bound for low flows" (Smakhtin, 2001: 153).

Streamflow intermittency generally poses a problem, as in grid cells where the threshold (in this case Q50) is zero in a particular calendar month, droughts are never identified in this month. To overcome this problem, ~~CQDI1_Q50~~CQDI1(Q50) allows an existing drought to continue during months with Q50=0, but only if Q in the respective month is also zero. In months where Q50 is zero, but Q exceeds zero, the drought event ends. This approach implies that a drought event can be prolonged, but never begin in calendar months with Q50=0.

~~**CQDI1_Q80**~~**CQDI1(Q80)** was calculated in the same manner as ~~CQDI1_Q50~~CQDI1(Q50), but ~~however,~~ using ~~the empirical percentile~~ Q80 per calendar month as threshold. With Q80, a deficit is computed in 20% of the 30 calendar months. ~~, which is the monthly streamflow value that is exceeded in 80% of all 30 calendar months.~~ Q80 was computed in R using the quantile function with the default quantile algorithm, such that Q80 is a streamflow value slightly higher than the sixth lowest calendar month streamflow. Daily or monthly Q80 is often used as a threshold for defining the onset and termination of a streamflow deficit period (van Huijgevoort et al., 2014; van Loon et al., 2014; Heudorfer and Stahl, 2017; Laaha et al., 2017), but the selected threshold should represent local water requirements (including environmental flow) (Cammalleri et al., 2016a).

~~**CQDI1_Q80_HS**~~**CQDI1(Q80-HS)** is a variant of ~~CQDI1_Q80~~CQDI1(Q80) suitable in intermittent and highly seasonal (HS) streamflow regimes where people strongly rely on water storage in man-made reservoirs that needs to be replenished by streamflow. It allows an existing drought to continue in any month where Q80 is zero also if the current streamflow Q exceeds zero. However, the cumulative deficit is reduced by any streamflow surplus over the calendar month Q80. The rationale behind this approach is that streamflow during low-flow months (calendar months where Q80 is zero) is not relevant for people relying on large reservoirs. Below-normal water storages can only marginally be replenished during a low-flow period, and hence drought severity should remain at the level of the preceding high-flow period. Like ~~CQDI1_Q80~~CQDI1(Q80), a drought can be prolonged but never begin, in months with Q80=0.

CQDI6(Q80) is computed like CQDI1(Q80) but applying an averaging period of six months. The indicator is suitable in regions with access to large reservoirs. In each month, the streamflow deficit is computed by subtracting the average streamflow of the preceding six months (including the current month) from the long-term Q80 of the same six months during the reference period.

2.~~2~~3.~~4~~ 3 Empirical percentiles EP1 and cumulative empirical percentiles CEP1(20%)

Empirical streamflow percentiles **EP1** were computed per calendar month following Eq. (2) with an averaging period of one month. EP1 expresses the frequency of non-exceedance, while the inverse is the return period, in years, with ~~,~~

$$EP1 = rank(Q)/n \quad (2)$$

where rank(Q) is the rank of a streamflow value of a certain calendar month and n is the sample size, i.e., the number of years in the reference period.

Rank 1 was assigned to the smallest streamflow value. If a sample contained several months with the same streamflow value, the largest rank among these months was assigned to the tied streamflow values. For a calendar month comprising, for instance, 26 out of 30 months with zero streamflow, a value of EP1=26/30 would be assigned to the respective 26 months corresponding to a return period of 1.2 years. This method slightly adjusts the approach by Tijdeman et al. (2020), who used the average rank among the tied values. In the given example, this would result in EP1=0.45 and a return period of 2.2 years for the first 26 values. In this study, we chose the largest EP1 for tied values to reflect that frequent streamflow values have a high frequency of non-exceedance and a low return period assuming that people and the ecosystem are habituated to more frequent values including zero streamflow.

~~The cumulative percentile-based anomaly~~ **CEP1(20%)** is the ~~-~~cumulative percentile-based deficit~~anomaly was computed in a similar way to CQDI1-Q80~~. The monthly percentile deficit is computed by subtracting the current streamflow percentile from a selected percentile threshold (Eq. 3). In this study, a deficit is computed ~~using the 20th percentile (the value that is exceeded in~~ for the six lowest calendar month values (20% out of 30 values). Consequently, the selected threshold percentile is slightly higher than 20% depending on the sample size (22.7% in this study with a sample size of 30 and 22% for a sample size of 40). Monthly percentile deficits are accumulated for all drought months to obtain severity. ~~8 out of 10 months) of the respective 30 EP1 values as threshold per calendar month. Moreover~~~~Similar to~~Like CQDI1(Q80), CEP1(20%) allows an existing drought event to continue during months where both Q80 and the current streamflow are zero. T~~-~~he 2mc is also applied. Hence, CEP1(20%) identifies the same drought months as CQDI1(Q80). The percentile deficit is computed as

$$\text{percentile deficit}_{m,y} = P20 - EP1_{m,y} \quad \text{(for } EP1_{m,y} < P20\text{)} \quad (3)$$

with m = month, y = year, ~~percentile deficit$_{m,y}$ = current monthly deficit,~~ and $EP1_{m,y}$ = current empirical streamflow percentile. With a sample size of 30 calendar month values, the percentile threshold $P20_m$ is 22.7% such that 20% of all calendar months are identified as drought months.

2.~~2~~3.~~5~~ 4 Relative deviation from mean conditions RQDI1, RQDI12 and cumulative CRQDI1(-50%)

RQDI1 is the relative deviation of monthly streamflow from mean calendar month streamflow (MMQ) in percent. In each month, it is calculated as the difference between monthly streamflow and the respective MMQ, which is then divided by MMQ.

-**RQDI12** is the relative deviation of mean streamflow during the preceding 12 months (in km³ month^{-1}) from mean annual streamflow (in km³ month^{-1}) during the reference period. In this study, RQDI12 is only assessed for ~~two~~ selected gauging stations (~~Fig. 2 and Sect. 3.3~~add reference supplement), but not at the global scale.

The cumulative relative deviation **CRQDI1(-50%)** is computed ~~with~~ using a threshold of RQDI1=-50% and applying the 2mc (Sect. 2.3.2). ~~was derived like CQDI Q80.~~ Months with MMQ=0, where the relative deviation is not computable, ~~were~~ are defined to end a drought event assuming that people are habituated to zero streamflow in this month~~.~~. The percent deficit ~~in each month~~ is computed as~~:~~

$$\text{percent deficit}_{m,y} = -50\% - \text{RQDI1}_{m,y} \quad \text{(for RQDI1} < 50\%) \quad (4)$$

with m = month, y = year, and $\text{RQDI1}_{m,y}$ = current relative streamflow deviation in percent.

2.~~2~~3.~~6~~ 5 Water deficit indicators CQDI1(-WUs) and CQDI1-(WUs-EFR)

The water deficit indicators ~~CQDI1-WUs~~CQDI1(WUs) and ~~CQDI1-WUs-EFR~~CQDI1(WUs-EFR) are computed like ~~CQDI1-Q80~~CQDI1(Q80) but using as thresholds mean monthly potential surface water abstraction WUs, and WUs plus environmental flow requirement (EFR), respectively. Following Richter et al. (2012), EFR is assumed to be 80% of mean monthly naturalized streamflow Qnat per calendar month such that 12 EFR values are obtained per grid cell. WUs is the simulated water demand (potential water abstractions from surface water bodies) and not the actual water abstractions (Müller Schmied et al., 2021), but both values are similar in most grid cells. The satisfied (or actual) water use is not suitable ~~to~~ for identifying periods of water deficit because it decreases along with water availability during drought. Cumulative deficits are normalized by mean annual streamflow. The indicators were not computed in grid cells where mean annual surface water demand in the reference period is zero (approx. 9% of all grid cells excluding Greenland). For CQDI1(WUs), the water deficit ~~in each month~~ is computed as~~:~~

$$\text{water deficit}_{m,y} = \text{WUs}_m - Q_{m,y} \quad \text{(for } Q_{m,y} < \text{WUs}_m) \quad (5)$$

with m = month, y = year, WUs_m = mean potential surface water abstraction per calendar month, and $Q_{m,y}$ = current streamflow.

For CQDI1(WUs-EFR), the water deficit in each month is computed as:

$$\text{water deficit}_{m,y} = \text{WUs}_m + \text{EFR}_m - Q_{m,y} \quad \text{(for } Q_{m,y} < (\text{WUs}_m + \text{EFR}_m)) \quad (6)$$

with m = month, y = year, WUs_m = mean potential surface water abstraction per calendar month, EFR_m = 80% of mean monthly naturalized streamflow Qnat per calendar month, and $Q_{m,y}$ = current streamflow.

2.~~3~~ 4 Probability of non-exceedance and return period of drought events of a certain severity

Following the approach of Cammalleri et al. (2016a) to compute the low-flow index LFI, the probability of drought events of a certain severity was computed for six cumulative indicators, CEP1(20%), four CQDI1 variants (thresholds Q50, Q80, WUs, and WUs+EFR) and CRQDI1(-50%). First, the partial duration series of drought events was derived based on the severities of all drought events of the reference period. Grid cells with less than six drought events were excluded. The exponential cumulative distribution function proposed in Cammalleri et al. (2016a) was used to estimate the probability of non-exceedance p of a certain cumulative streamflow deficit:

$$p(S_i; \lambda) = 1 - e^{-\lambda S_i} \qquad \text{(with } S_i > 0\text{)} \qquad (3\text{-}7)$$

where the variable S_i is the severity of drought event i, as quantified by a cumulative indicator, and the parameter λ is the inverse of the mean of the severities of all completed drought events. For instance, a value of p=0.7 in ~~March 2002~~a certain month denotes that, if the drought event ended in this month~~March 2002~~, its severity would be larger than the severity of 70% of the drought events in the reference period. Different from LFI, which is based on daily streamflow data, time series of monthly streamflow were used for all indicators and the 2mc (see Sect. 2.~~2~~3.~~3~~2) was applied. Since p was computed for each month of the reference period, it describes the non-exceedance probability (or rather frequency) of both completed drought events and continuing droughts. Following Sharma and Panu (2015) and Beguería (2005), the return period T_{ri} of a drought event with severity S_i is computed as

$$T_{ri} = \frac{1}{\theta(1-p(S_i))} \qquad (8)$$

where θ is the average number of drought events per year during the reference period.

3 Proposed systematic approach for selecting and computing SDHIs

Wilhite and Glantz (1985) suggested distinguishing between a conceptual and an operational drought definition, with the former referring to the general qualitative concept of drought and the latter allowing for a quantitative drought characterization including onset, severity, termination, and spatial extent. In the following Sect. 3.1, aspects that relate to the conceptual drought definition are discussed comprising the description of the targeted drought risk and the system at risk. In particular, assumptions about the habituation of the system at risk to the streamflow regime are discussed, an aspect that is currently not taken into account or not made explicit in drought hazard studies. ~~In order to~~To translate these conceptual definitions into operational drought hazard indicators, a new classification system for hazard indicators is proposed in Sect. 3.2. The new systematic approach is illustrated in Sect. 4 using modeled SDHIs at the global scale as well as observation-based SDHIs

~~3.3Supplementary material for selected SDHIs using streamflow observations~~ at ~~two~~ four gauging stations with different streamflow regimes and different assumed levels of ~~habituation~~vulnerability.

3.1 Assumptions about habituation inherent in drought hazard indicators

The selection of drought hazard indicators for a DEWS requires a clear definition of "the risk of what for whom". Drought hazard indicators are risk system-specific (Blauhut et al., 2021), and there is not one that fits all. Drought is usually conceptualized as anomaly ("less water than normal") and/or deficit ("less water than needed"). Consequently, the selection of an indicator requires a definition, often based on assumptions, about "what is normal or needed", i.e., to what the risk system is habituated to. ~~The choice of drought hazard indicators implies assumptions about the habituation of the system at risk.~~ In the case of streamflow, people and ecosystems are assumed to have adapted to certain characteristics of the flow regime. For example, if drought indicators are computed based on the calendar month-specific distribution of streamflow values, it is implicitly assumed that ~~people and ecosystems are~~the risk system has adapted to the seasonality of streamflow. ~~In case of SSI1, it is further assumed that people and ecosystems are adapted to a certain degree of interannual variability, e.g., to the low streamflow that is only exceeded in 1 out of 5 years.~~ But also temporally constant thresholds, which have traditionally been used to define hydrological droughts (Stahl et al., 2020), are suitable for certain systems, e.g., for computing drought risk for electricity generation by thermal power plants, which require a certain minimum streamflow for operation.

~~When conceptualizing or selecting a hazard indicator for a specific drought risk, these assumptions on habituation of the system at risk should be made explicit.~~ At the global scale, it is unknown to which streamflow characteristics different risk systems such as drinking water supply, irrigation water supply, hydropower production, and the river ecosystem are accustomed. Therefore, the ~~twelve~~ eleven global-scale drought hazard indicators analyzed in this study (Table 1) cover different types of habituation, including the habituation to a certain degree of interannual variability of streamflow, to streamflow seasonality, to a certain reduction from mean calendar month or mean annual streamflow, and to being able to fulfill the demand for surface water abstractions and environmental flow. It is up to the user of a large-scale DEWS, who understands the local risk system-specific habituation to reduced water availability, to select the hazard indicator that is appropriate for the risk system of interest. ~~Table 1 lists the indicators according to this classification together with unique characteristics relevant for streamflow drought risk assessments.~~

Table 1: Characteristics of ~~conventional~~ SDHIs suitable for global-scale assessments, classified according to inherent assumptions about habituation of people or other biota. The general terms "a certain degree" or "a certain reduction" in the first column are specified in a drought assessment by selected thresholds for drought definition.

Assumed habituation and suitable indicator *People or other biota accustomed to*		**Characteristics**
a certain degree of inter-annual variability	SSI12, ~~/~~EP12[1], CQDI1 ~~(~~Q80-HS), CQDI6(Q80)	Suitable for quantifying 1) risk for human water supply in regions with large man-made reservoirs or lakes that buffer seasonal streamflow deficits as well as 2) risk for large lake and wetland ecosystems.
	~~SPI12~~	~~If used as proxy for streamflow drought hazard, assumptions about habituation are the same as for SSI1.~~ ~~Processes in altered flow regimes cannot be characterized.~~
~~seasonality and a certain degree of interannual variability~~	~~SPEI12~~	~~Same characteristics as SPI12; better proxy for streamflow drought hazard as it takes into account the impact of increased potential evapotranspiration on drought.~~
seasonality and a certain degree of interannual variability	SSI1~~/~~, EP1~~/~~, ~~CQDI1-Q80~~CQDI1(Q80)	Suitable risk for human water supply and for risk for river ecosystems in regions without access to reservoirs. Streamflow drought hazard might be underestimated in regions with high vulnerability and interannual variability.
seasonality	and median calendar month streamflow ~~CQDI1-Q50~~CQDI1(Q50)	Using such a high threshold (median of calendar monthly streamflow) can be beneficial in highly vulnerable regions where people cannot even cope with small reductions ~~in~~ from median calendar month~~ly~~ streamflow.
	being able to fulfill demand for surface water abstractions ~~CQDI1-WUs~~CQDI1(WUs)	The system at risk is accustomed to the seasonality of human water demand (WUs). People are used to being able to fulfill human water demand. The health of river ecosystems is not taken into account. An indicator of water deficit rather than drought hazard.
	being able to fulfill demand for surface water abstractions and environmental flow ~~CQDI1-WUs-EFR~~CQDI1(WUs-EFR)	The system at risk is accustomed to the seasonality of human water demand (WUs) and to the seasonality of environmental flow requirements (EFR). Alternative 1: EFR based on Qant[2]: The river ecosystem has adjusted to the altered flow regime over the last decades, which is considered the "new normal status". Alternative 2: EFR based on Qnat[2]: the natural flow regime is the aspired status.
	a certain reduction from mean calendar month streamflow RQDI1	Suitable in highly vulnerable regions where people cannot even cope with small reductions from mean calendar month streamflow. ~~Suitable in study regions without large surface water storages.~~ Drought hazard might be overestimated in regions with low vulnerability and low interannual variability.
a certain reduction from mean annual streamflow	RQDI12	Suitable in study regions with large man-made reservoirs or lakes, which buffer seasonal streamflow deficits. Drought hazard might be overestimated in regions with low vulnerability and interannual variability.

temporally constant minimum streamflow	Not included in this study	Identifies drought hazard whenever water availability drops beneath a certain level (e.g., water intake for cooling of thermal power plants has to be reduced). Identifies no drought in the wet season.

[1] EP12: Empirical streamflow percentile with an averaging period of 12 months (not analyzed in this study)

[2] Qant, Qnat: Modeled anthropogenic streamflow altered by human water use and man-made reservoirs (Qant) and naturalized modeled streamflow (Qnat)

~~In hydrology, flow duration curves showing the fraction of the time that a certain streamflow is exceeded (expressed as percentiles) are a widely used method to assess the low flow regime (Smakhtin, 2001).~~ Percentile-based indicators including empirical streamflow percentiles, standardized indicators, and TLM indicators with a low streamflow percentile as threshold are often applied in DEWS (Bachmair et al., 2016; Cammalleri et al., 2016a). They are perceived as statistically consistent across different temporal and spatial scales, indicating the rarity of the event (Steinemann et al., 2015; WMO and GWP, 2016). ~~Indicators of less than normal water availability such as "percent of normal precipitation" appear to be less preferred as time periods with the same indicator value have different probabilities of occurrence in different regions and thus not the same rarity (Steinemann et al., 2015). However, according to Kumar et al., (2009)., percent deviations from mean precipitation have been used to assess drought intensity in India, South Africa, and Poland. Kumar et al. (2009) compared percent precipitation deviations and SPI in two districts in India, one in a humid region with high mean precipitation and low interannual variability and the other in a semi-arid region with low mean precipitation and higher interannual variability. Based on a 39-year record of observed monthly precipitation they showed that much higher percent deviations occurred, in the case of SPI = -1, in the low precipitation district than in the high precipitation district, e.g., 70% and 30%, respectively. Consequently, drought hazard may be underestimated with SPI in the low precipitation district. For example, yield loss is more closely related to percent of normal precipitation than to the rarity of the low precipitation event, as crop yield depends on actual evapotranspiration in percent of PET, which decreases with precipitation (Siebert and Döll, 2010).~~

~~Application~~Utilization of percentile-based indicators (e.g., SSI12, SSI1, and CQDI1 ~~(Q80)CQDI1(Q80)~~ in Table 1) implies that people in different climate regions and social systems are equally habituated to a certain interannual variability, which is most likely not the case. ~~Similar to the example above from Kumar et al. (2009), t~~The 20th streamflow percentile (or SSI1 = -0.84) would correspond to a low relative streamflow deviation (e.g. -20%) in a humid region (low interannual variability) compared to a higher deviation (e.g. -50%) in a semi-arid region (high interannual variability). Hence, ~~although~~ percentile-based indicators ~~have the advantage of being spatially comparable in terms of drought frequency, they~~ might underestimate streamflow drought hazard in semi-arid areas where people (and ecosystems, albeit possibly to a lower degree) are often more vulnerable to reductions in water availability. Regions with high interannual variability are depicted in Fig. A1b. Here, drought hazard indicators that quantify relative deviations from the long-term mean or median (RQDI1, RQDI12 in Table 1) or TLM indicators with higher percentiles as threshold (CQDI1(Q50) in Table 1) might be better suited to define drought conditions. Such indicators appear to be less preferred as periods with the same indicator value have different probabilities of occurrence in different regions and thus not the same rarity (Steinemann et al., 2015). ~~These include percent deviations from~~

~~mean streamflow (RQDI1, RQDI12 in Table 1) or TLM indicators with higher percentiles as threshold (CQDI1-Q50~~CQDI1(Q50) ~~in Table 1).~~ Contrastingly, river ecosystems are, in the ideal case, perfectly adjusted to interannual variability of streamflow such that percentile-based drought hazard indicators are often suitable for drought risk assessment for river ecosystems. In conclusion, percentile-based hazard indicators and relative deviations from the long-term mean or median should be used complementarily in large-scale DEWS to cover different drought risks.

T~~Another important characteristic of drought hazard indicators is t~~he selected averaging period ~~that~~ defines whether people are habituated to the annual or seasonal flow regime. One can assume that river ecosystems are generally accustomed to seasonality. Therefore, indicators with a short averaging period of, for example, one month (EP1, SSI1, RQDI1 and CQDI1 variants in Table 1) are appropriate for quantifying drought hazard for river ecosystems. Furthermore, short averaging periods are suitable in regions where farmers and other water users do not have access to large water storages such as reservoirs, lakes, or groundwater (either due to missing infrastructure or due to water use restrictions). As these users abstract water directly from the stream, they are very vulnerable to seasonal (monthly) streamflow deficits. Indicators with longer averaging periods (SSI12, ~~and~~ RQDI12), on the other hand, are suitable in regions with large man-made reservoirs, which are usually replenished during the wet season such that streamflow deficits during the low-flow months are irrelevant. People in these regions are therefore only vulnerable to either interannual variability (SSI12) or mean annual conditions (RQDI12), but not to seasonality. Certainly, other averaging periods may be suitable depending on the region-specific storage capacity.

Since volume-based indicators (TLM indicators) are also important components in water resources management (van Loon, 2015), the ~~we propose the volume based~~ indicators ~~CQDI1-Q80-HS~~CQDI1(Q80-HS) and CQDI6(Q80) are assessed as ~~an~~ alternatives for SSI12 and RQDI12 (or rather the cumulated variants CSSI12 and CRQDI12) in regions with highly seasonal ~~(HS)~~ streamflow regimes (Fig. A1a) and large reservoirs.

~~If water users need streamflow to fill a reservoir, streamflow availability during the dry season would be of (almost) no interest to the risk takers/water users. For them, it would be worse to have less water than normal during two consecutive wet seasons even if there is slightly more water than normal in the dry season (as the amount of this water is very small compared to the water produced in the wet season). With CQDI1-Q80-HS~~CQDI1(Q80-HS), ~~an existing drought is allowed to continue during a pre-defined low-flow period, namely months where the calendar month Q80 is zero, even if streamflow exceeds zero (Sect. 2.2.3). Consequently, in case of two consecutive wet season droughts the streamflow deficit continues to accumulate resulting in a higher drought severity for the pooled drought event than for the two single wet season droughts. This most likely reflects the perceived hazard of such a situation better. Regions with highly seasonal streamflow where the application of CQDI-Q80-HS is meaningful are depicted in Fig. A1a.~~

~~Obviously, if SPI12 and SPEI12 are used to assess meteorological drought hazard, people and ecosystems are assumed to be habituated to the interannual variability, but not the seasonality, of P and P-PET. However, when they are used as proxies to identify streamflow drought hazard, they should ideally correspond to the temporal development of SSI1. Their performance would be assessed by comparing the goodness of fit between time series of SPI12 or SPEI12 with SSI1. In this case, assumptions about the habituation inherent in SPI12 and SPEI12 refer to the streamflow regime and not to time series of P~~

and P-PET. Accordingly, SPI12 and SPEI12 fall into the same category as SSI1 in Table 1 (interannual variability and seasonality). The suitability of different averaging periods for SPI for describing streamflow drought hazard is discussed in Sect. 4.2.5.

For water managers, the status of the actual water deficit in terms of unsatisfied water demand might be as informative as the status of streamflow anomaly. Drought hazard is generally defined as a climate-induced anomaly, i.e., a period of below-normal water availability (McKee et al., 1993; van Lanen, 2006; van Loon, 2015). This concept can be broadened by assuming that a drought only occurs if the anomaly coincides with a water deficit for people or ecosystems (Cammalleri et al., 2016b; Popat and Döll, 2021; Wilhite and Glantz, 1985). This concept is not new and several definitions were already summarized in Wilhite and Glantz, 1985). , e.g., drought is a "period during which streamflows are inadequate to supply established uses under a given water management system" (Linsley et al., 1975 in Wilhite and Glantz, 1985: 115). Nevertheless, only a few studies exist where the combination of anomaly and deficit was translated into drought hazard indicators for soil moisture (Palmer, 1965; Cammalleri et al., 2016b; Popat and Döll, 2021) and streamflow (Popat and Döll, 2021). In the present study, the water deficit aspect of drought is represented by the indicators CQDI1-WUsCQDI1(WUs) and CQDI1-WUs-EFRCQDI1(WuUs-EFR) where surface water demand is taken as the threshold (Table 1). Application of these indicators implies that the system at risk is habituated to the satisfaction of seasonal water demand. While CQDI1-WUsCQDI1(WuUs) neglects the water requirements of the ecosystem, CQDI1-WUs-EFRCQDI1(WuUs-EFR) assumes that the river ecosystem is habituated to the seasonality and magnitude of natural streamflow. As EFR might never be fulfilled during the investigation period in the case of streamflow regimes that are strongly altered streamflow regimesby human water abstractions and man-made reservoirs, Qnat in the EFR computation can be replaced by Qant implying . This implies the assumption that the river ecosystem has already adapted to the altered streamflow conditions (Table 1). Figure A1c shows regions where human water demand is high compared to available streamflow and where a drought hazard due to unsatisfied human surface water demand is likely.

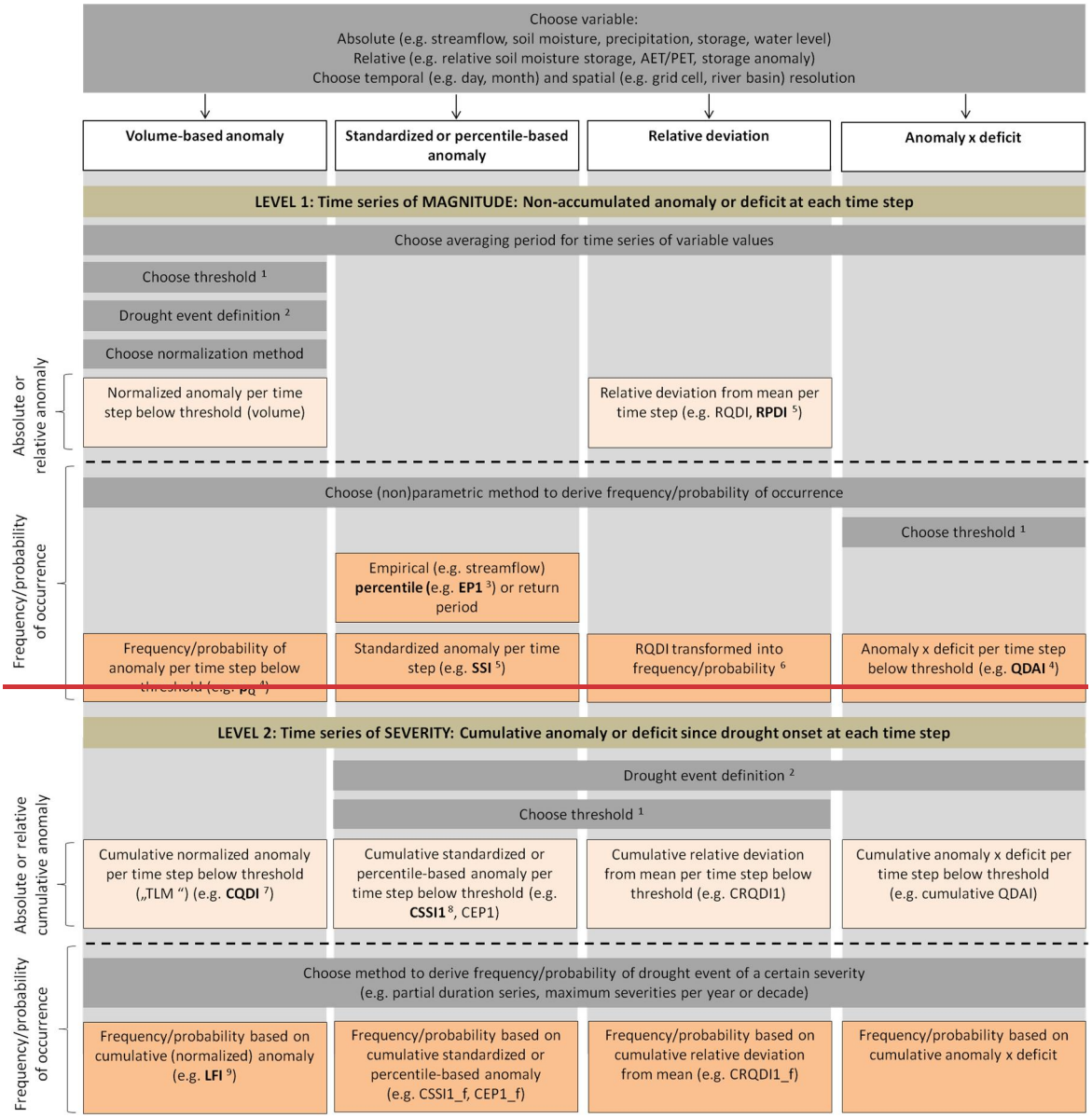
3.2 Levels of drought characterization

Drought risk indicators, and thus drought hazard, exposure and vulnerability indicators, should be designed specifically for the targeted risk (Lloyd-Hughes, 2014; Spinoni et al., 2018; Hagenlocher et al., 2019). As an example, the risk of "not being able to provide enough water to fulfill the customers' water demand during times of lower water availability than normal" constitutes a drought risk for water supply companies. If the water supply source is a river, a suitable drought hazard indicator should be based on streamflow data or on water storage in the upstream reservoir. However, a very large number of drought hazard indicators has been proposed and applied by experts without stringent consideration of the targeted risks (Wilhite and Glantz, 1985; Lloyd-Hughes, 2014). To identify a suitable risk-specific hazard indicator, the first step is to decide which water flow or storage should be taken into account, e.g. precipitation, soil moisture or streamflow, as the determined drought hazards depend on the considered physical variable (Satoh et al., 2021). Even after determining the appropriate

~~risk-specific physical variable, e.g. streamflow, there is still a large choice of possible drought hazard indicators to quantify the occurrence or severity of streamflow droughts (WMO and GWP, 2016; Yihdego et al., 2019).~~

Translating conceptual drought definitions into operational, quantitative drought hazard indicators is not straightforward due to the complexity of the underlying natural processes and the large number of indicators and methods that can be applied. In the literature, there is agreement about which drought characteristics are relevant for operational applications comprising the temporal component (onset, termination, duration) and the spatial extent as well as drought magnitude and severity, from which other metrics such as intensity, return period, and frequency or probability of occurrence can be derived (van Lanen et al., 2017). We understand drought *magnitude* as an anomaly or deficit occurring within a pre-defined period and *severity* as the accumulated deficit between the magnitude and a selected threshold since the onset of drought, which is defined by water availability dropping below the threshold (van Lanen et al., 2017). However, the terms drought magnitude and severity, which represent different levels of drought characterization, are not applied consistently in the literature. The terms are not made explicit and are sometimes interchanged (Steinemann et al., 2015, Vidal et al., 2009; López-Moreno et al., 2009). In particular, the commonly accepted classification of SDHIs into threshold-based and standardized indicators (van Loon, 2015) is somewhat misleading, since the former represents time series of severity and the latter time series of magnitude.

To facilitate a better understanding of the informative value of SDHIs, we suggest a new indicator classification that includes four types of indicators and distinguishes severity from magnitude indicators (Fig. 1). The indicator types (columns in Fig. 1) include the volume-based anomaly, the standardized or percentile-based anomaly, and the relative deviation (Sect. 2.3)~~, all of which are described in the previous section~~. Deficit-anomaly indicators (last column in Fig. 1) combine an anomaly indicator with an indicator of the deficit with respect to optimal water availability (e.g., ~~. For example, Popat and Döll (2021)~~Popat and Döll, 2021). ~~combined the volume based anomaly indicator p_Q (Fig. 1) with an indicator of the streamflow deficit with respect to water demand to obtain the streamflow deficit anomaly indicator QDAI.~~ For each indicator type, two levels of drought characterization, drought magnitude (level 1) and severity (level 2), can be computed~~. Level 1 indicates the drought magnitude at each time step~~. ~~Time steps in drought analysis are usually months, but daily time steps may be used in drought monitoring systems (Cammalleri et al., 2016a). Time series of drought magnitude can be expressed as absolute (volume based) or relative anomaly or deviation or in terms of frequency or probability of occurrence. If magnitude indicators are cumulated since drought onset, severity indicators are obtained at level 2.~~ ~~The units of the four indicator types differ both at level 1 and 2, but at level 2, indicators can be directly compared when expressed in units of probability of non-exceedance (Fig. 1).~~



[1] Threshold to define anomaly or deficit based on, e.g. the same variable or a type of (human, plant, ecosystem) water demand and defined by, e.g. constant or seasonal values, percentiles or mean, temporally averaged over the averaging period of the variable or over a different time period (e.g. 31-day running mean of daily values in case of a daily averaging period)

[2] Methods to handle periods of no or low flow or storage; definitions for onset, termination, pooling of drought events

[3] e.g. Tijdeman et al. (2020)

[4] Döll and Popat (2021) (QDAI: streamflow deficit anomaly index; p_Q: streamflow drought probability index)

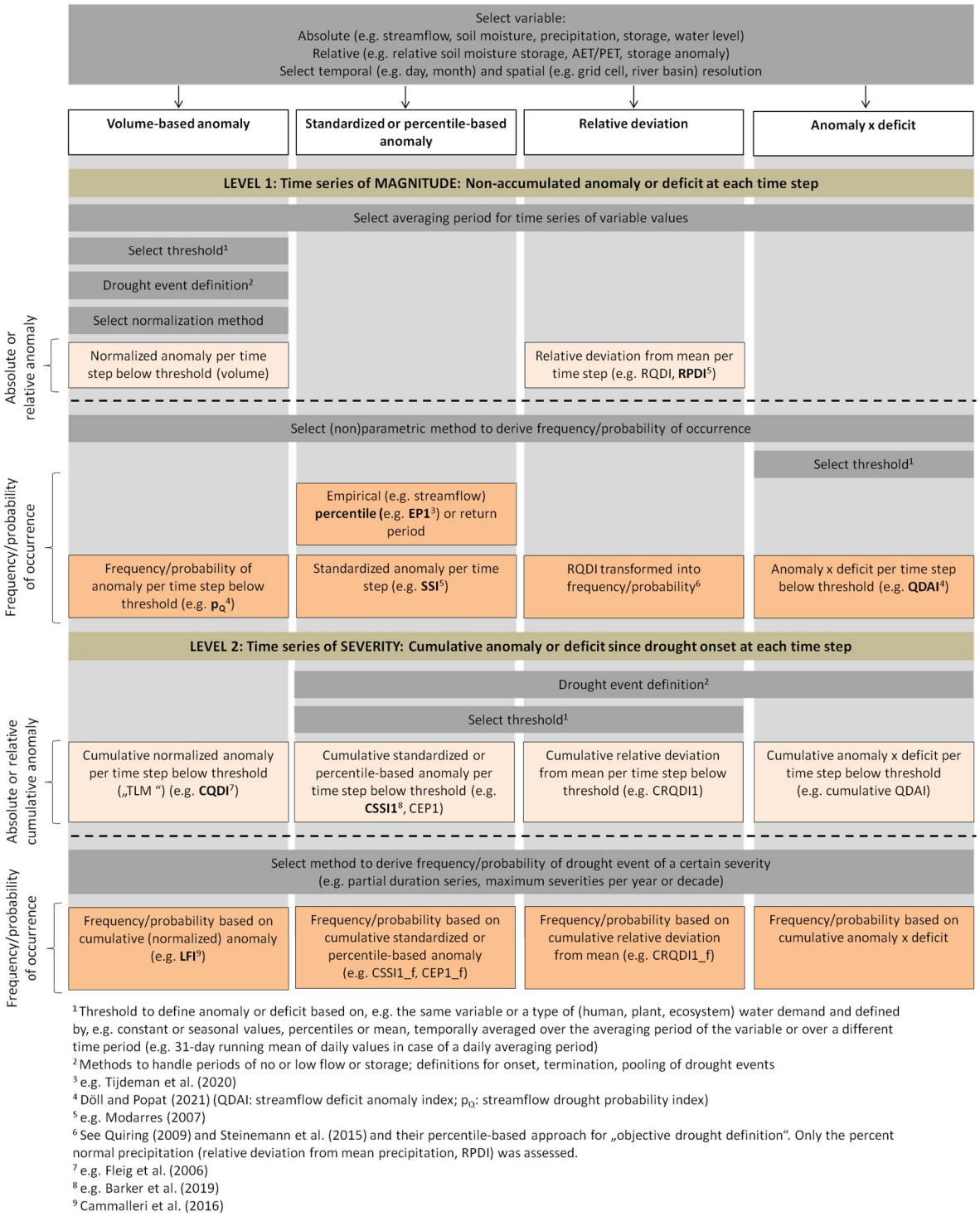
[5] e.g. Modarres (2007)

[6] See Quiring (2009) and Steinemann et al. (2015) and their percentile-based approach for „objective drought definition". Only the percent normal precipitation (relative deviation from mean precipitation, RPDI) was assessed.

[7] e.g. Fleig et al. (2006)

[8] e.g. Barker et al. (2019)

[9] Cammalleri et al. (2016)



[1] Threshold to define anomaly or deficit based on, e.g. the same variable or a type of (human, plant, ecosystem) water demand and defined by, e.g. constant or seasonal values, percentiles or mean, temporally averaged over the averaging period of the variable or over a different time period (e.g. 31-day running mean of daily values in case of a daily averaging period)

[2] Methods to handle periods of no or low flow or storage; definitions for onset, termination, pooling of drought events

[3] e.g. Tijdeman et al. (2020)

[4] Döll and Popat (2021) (QDAI: streamflow deficit anomaly index; p_Q: streamflow drought probability index)

[5] e.g. Modarres (2007)

[6] See Quiring (2009) and Steinemann et al. (2015) and their percentile-based approach for „objective drought definition". Only the percent normal precipitation (relative deviation from mean precipitation, RPDI) was assessed.

[7] e.g. Fleig et al. (2006)

[8] e.g. Barker et al. (2019)

[9] Cammalleri et al. (2016)

590

Figure 1: Classification system including four types of drought hazard indicators, indicating 1) magnitude of the drought at a certain time step as deficit and/or anomaly (level 1) or 2) severity of the drought event, i.e. the cumulative magnitude of drought since

drought onset (level 2). Both magnitude and severity can be expressed in terms of frequency/probability to compare the drought of interest to other droughts. The dark grey boxes indicate decisions that have to be made when computing the indicators, e.g. which averaging period is selected. Indicators in bold have already been applied in the literature. Assumptions about the habituation of people and ecosystems determine the selection of the type of indicator, the averaging period, and the threshold (see Table 1).

The dark grey boxes in Fig. 1 represent decisions ~~that have to be made~~ regarding time step length and averaging period, drought threshold and definition of drought events (minimum length of drought event, pooling of drought events). These decisions depend on the assumed habituation of people and ecosystems to certain streamflow conditions (Sect. 3.1 and Table 1). Beige and orange boxes contain indicators that are expressed in absolute or relative values and in frequency/probability of occurrence, respectively. Indicators applied in drought monitoring (CQDI1, low-flow index LFI, percentiles, SSI, RDPI) or in the literature (p_Q, cumulative SSI, streamflow deficit anomaly indicator QDAI) are written in bold. The units of the four indicator types differ both at level 1 and 2, but ~~at level 2,~~ indicators can be directly compared when expressed in units of probability (or frequency) of non-exceedance ~~(Fig. 1)~~.

Figure 1 shows that ~~the specific~~ drought hazard indicators ~~represent different levels of drought characterization (magnitude and severity) and that those~~ pertaining to one of the four indicator types can be transformed between level 1 (magnitude) and level 2 (severity) while still sharing the type-specific conceptual drought definition. Furthermore, the classification system clarifies that each indicator type requires a threshold setting either at level 1 or 2. Hence, the term "threshold-based" applies to any indicator of drought severity and it is therefore not a suitable criterion for distinguishing types of indicators.

The ~~classification~~ differentiation of indicator types can be ambiguous. For instance, standardized and percentile-based anomaly indicators are subsumed in Fig. 1 (column 2), although there is a minor conceptual difference between them as highlighted by Tijdeman et al. (2020). While standardized indicators show the non-exceedance probability enabling extrapolation, empirical percentiles represent the historical non-exceedance frequency within the boundaries of observations. We account for this aspect by including the terms frequency and probability in Fig. 1.~~.~~

~~On the other hand, v~~Volume-based and standardized or percentile-based anomaly indicators~~, on the other hand,~~ are presented as different indicator types, although they can be based on the same conceptual drought definition if equivalent thresholds are applied. If Q80 is used as threshold for CQDI1 and -0.84 for cumulative SSI1 (corresponding to the 20th percentile for cumulative EP1 and a return period of 5 years), both indicators capture the same drought signal. Differences between the drought signals are then attributable to the computational methods for the standardization of streamflow. Analyzing the sensitivity of SSI1 to different parametric and nonparametric standardization methods in European river basins, Tijdeman et al. (2020) ~~Tijdeman et al. (2020)~~ revealed considerable differences in computed SSI1 among seven probability distributions (and two fitting methods) and five non-parametric methods.

A major difference between ~~CQDI~~volume-based and standardized indicators is that the former ~~drought severity can be expressed in volume of "missing" water and thus in absolute rather than relative values, which is often more informative in water resources management (van Loon, 2015). Although both indicator types capture the same drought signal (see above),~~

~~the relative levels of drought severity among the drought events during the reference period differ. Volume-based indicators~~ detect absolute drought deficits and ~~standardized or percentile-based indicators~~the latter relative drought deficits. This can result in different frequency values for the same drought event.~~As an example, a monthly deficit volume of 1% of mean annual streamflow represents a larger deviation from median streamflow in a low-flow month compared to a high-flow month. Consequently, differences between the two indicator types can be large when severities of specific drought events during the reference period are compared. This difference is illustrated in Sect. 4.2.4 and Fig. 8.~~

~~3.3 Illustration of habituation-based classification approach~~

~~The relation among ten out of the twelve indicators in Table 1 was assessed for time series of monthly streamflow observed at two GRDC gauging stations with different streamflow regimes. The Little Colorado River near Cameron in the United States (Fig. 2, left) was selected because it is the only station among 220 GRDC gauging stations worldwide with continuous monthly streamflow observations between 1986 and 2015 that has a visible impact of the HS method (Sect. 2.2.3). It is characterized by comparably low mean annual streamflow (MAQ) (ca. 5 $m^3 s^{-1}$) and high interannual and seasonal variability. Q80 is zero in May, June, and November. The Danube River at Hofkirchen in Germany (Fig. 2, right) was selected due to the different streamflow regime (much higher MAQ of 640 $m^3 s^{-1}$, lower seasonal and interannual variability) and due to the fact that the 2003 European drought can be used as a benchmark for the assessment. Since both stations are situated in river basins with an assumed low vulnerability to drought by global comparison, the indicator CQDI1_Q50, suitable in highly vulnerable regions, was not considered. We used observations instead of WaterGAP modelling result to exclude model uncertainties. Only mean monthly WUs used in CQDI1_WUs and CQDI1_WUs_EFR is based on WaterGAP model output. Different from the description in Sect. 2.2.6, EFR in CQDI1_WUs_EFR is computed as 80% of observed mean monthly Qant and not Qnat.~~

~~In Fig. 2a, streamflow anomalies below Q80 are highlighted in orange. Fig. 2b depicts time series of drought severity according to different CQDI1 Q80 variants. First, the effect of the 2mc (Sect. 2.2.3) can be deducted by comparing CQDI1 Q80 (without 2mc) and CQDI1 Q80. Applying the 2mc, several one month droughts are excluded at both stations and two drought events are pooled into one in 1996 (Little Colorado River) and 2003 (Danube River). Comparing CQDI1 Q80 HS and CQDI Q80, the HS method (Sect. 2.2.3) can either lead to the mere prolongation of drought events (for example in 1990 and 1991 at the Little Colorado River) or to the pooling of two or more wet season droughts into one drought event (not identified at the two stations). When computing the frequency distribution of drought severity, there would be no difference between CQDI1 Q80 and CQDI1 Q80 HS in case of drought prolongation. In contrast, the pooling of two or more wet season droughts into one drought event does change the frequency distribution of drought severity. Whether the assumptions about habituation inherent in the 2mc and the HS method adequately reflect the perceived drought hazard can only be answered with regional knowledge about the vulnerability of the system at risk. At the Danube station, the 2mc certainly leads to the realistic extension of the drought event in 2003 until the end of the year.~~

~~For better comparison with the CQDI1 Q80 variants, only z scores below -0.84, equivalent to Q80, are shown for the standardized indicators (Fig. 2c). Since fitting of the gamma distribution was rejected for the Little Colorado River station~~

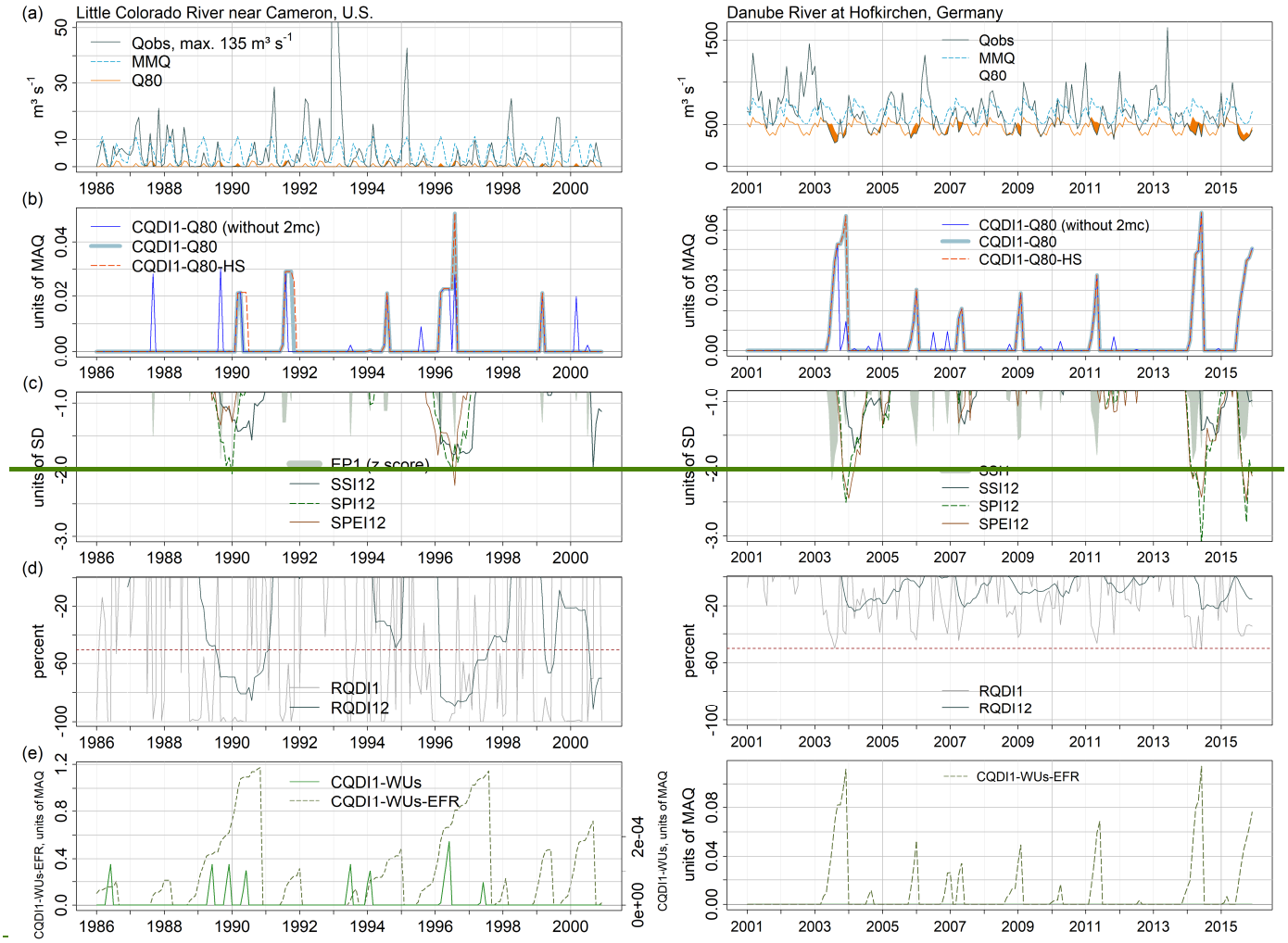
based on the KS test (Sect. 2.2.2), the indicator EP1 was computed instead of SSI1 and transformed into z scores. EP1 and SSI1, respectively, capture the same drought signal as CQDI1 Q80 (without 2mc) at both stations. Nevertheless, the former indicate the drought magnitude during the month of interest only, while CQDI1 Q80 indicates the severity since the beginning of the drought event. For instance, drought severity of the drought event in 2014 (Danube station) exceeds the value in 2011 by a factor of almost 2. The maximum drought magnitude for both events, however, is very similar according to SSI1. Cammalleri et al. (2016a: 356) aptly write that standardized indicators such as SPI and SSI cannot reproduce the "conceptual mechanism behind the evolution of a drought event as a phenomenon that is derived from a continuous hydrological quantity with daily values that are strongly dependent on the antecedent status". We argue though that this is just due to the fact that SSI1, as well as EP1, indicate drought magnitude and could be converted into the severity indicators CSSI1 and CEP1, respectively (see Fig. 1), which take into account the antecedent status of streamflow (Fig. 7f).

With SSI12 (Fig. 2c), deviations from normal conditions are smoothed over the preceding 12 months making the indicator suitable for identifying reservoir drought hazard. Streamflow drought events at both stations are indicated by SSI12 with a delay of several months with regard to drought onset and termination (2003 and 2014, Danube River) or drought termination only (1996, Little Colorado River). The correspondence between the meteorological indicators SPI12 and SPEI12 (Fig. 2c) and the hydrological indicators (SSI1 or EP1 and CQDI1 Q80 variants) is low at both stations. For most streamflow drought events, the averaging period of 12 months for the meteorological variables leads to excessive delays in the signal. Many short drought signals are not detected at all. Performance of SPI12 and SPEI12 is equally low at both stations. Hence, drought propagation through the hydrological cycle is faster than estimated by SPI12 and SPEI12. This is also supported by the sensitivity analysis of SPI averaging periods in Sect. 4.2.5. At both stations in Fig. 2, an averaging period of 3 months resulted in the highest correlation between SPI and observed SSI1.

RQDI (Fig. 2d) indicates the magnitude of streamflow drought hazard under the assumption that the system at risk is habituated to mean monthly streamflow but not to interannual variability. Due to the high interannual variability at the Little Colorado River with a few high flow years that considerably increase mean monthly streamflow, RQDI1 and RQDI12 are often below -50%. The strong RQDI1 signal is very different from the EP1 with minimum RQDI1 values below -90% corresponding to EP1 (z score) between -1.8 and +0.6. At the Danube station, the threshold -50% is only reached twice during the drought years 2003 and 2014 with SSI1 below -1.65. As discussed in Sect. 3.1, it is unknown at the global scale to which streamflow characteristics people and other biota are accustomed to, but Fig. 2 visualizes that SSI may underestimate the drought hazard in semi-arid regions. At the same time, RQDI probably overestimates drought hazard in regions where people are well accustomed to the interannual variability of streamflow.

The water demand deficit indicators CQDI1 WUs and CQDI1 WUs EFR (Fig. 2e) result in very different temporal patterns of drought severity as compared to the CQDI1 variants. While streamflow at the Little Colorado River is below Q80 mainly outside the low flow period (May-June and November-December), mean monthly WUs are highest in May and June, and consequently CQDI1 WU droughts often occur in these months. Drought severity according to CQDI1 WUs EFR is significantly higher and drought duration is much longer. EFR in Fig. 2e is computed as 80% of mean monthly observed Q.

~~Hence, it is assumed that the river ecosystem is adapted to the seasonality of streamflow, but it is negatively affected in years with very dry conditions. At the Little Colorado River, water deficits occur in 65% of all months during the depicted period and mainly stem from unsatisfied EFR. Application of CQDI1-WUs alone is not suitable to assess the current status of water deficit, as it does not consider the environmental component of water demand, but the indicator can be used complementarily to show the impact of human water demand on the total water deficit. At the Danube station, CQDI1-WUs is always zero, since WUs is only a small fraction of streamflow. Regions where human water demand is high as compared to supply include, e.g., the Mediterranean region, large parts of Turkey, India, and the western United States (Fig. A1c). Here, drought defined as water deficit due to high water demand is likely to occur. In these regions, CQDI-WUs and CQDI-WUs-EFR can indicate those months where human water use would have to decrease to alleviate drought burden on the river ecosystem.~~



~~Figure 2: Streamflow drought hazard based on observed streamflow during 1986-2015 at two WaterGAP calibration stations in the USA (Little Colorado River near Cameron, 1986-2000) (left) and Germany (Danube River, Hofkirchen, 2001-2015) (right): Monthly observed streamflow Qobs, mean monthly streamflow MMQ and Q80 (a); CQDI1-Q80 variants (b); SPI12, SPEI12, SSI1 or EP1 and SSI12 (c); RQDI1, RQDI12 (d); CQDI1-WUs and CQDI1-WUs-EFR (e). The cumulative indicators in (b) and (e) indicate drought severity; the non-accumulated indicators (c) and (d) indicate drought magnitude (see Fig. 1). In (a), periods where Qobs < Q80 are highlighted in orange. In (c, left), the z-score of EP1 is shown instead of SSI1, since gamma fitting was not possible. In (c), only the value range below -0.84 is shown. In (d), only negative relative deviations are depicted. 2mc: 2 months criterion. MAQ: mean annual streamflow. SD: standard deviation.~~

4 Similarities and ~~Quantification of global streamflow drought hazard by a global hydrological model~~discrepancies in SDHIs as quantified by a global hydrological model

The objective of this chapter is to identify which of the SDHIs presented in Table 1 can be meaningfully quantified at the global scale using WaterGAP 2.2d and which SDHIs are appropriate for monitoring different drought risks in large-scale DEWS. We emphasize that the objective is not a drought impact assessment, which is beyond the scope of this study. We want to show how the conceptual discrepancies and similarities between SDHIs (Sect. 3), which are of a general nature and apply to any month of the reference period, are translated into global-scale hazard indicators and how these indicators should be interpreted by end-users of a large-scale DEWS. The indicators are illustrated in global maps for two example months, July 2003 and July 2015, with known drought events in Europe. Following the classification of Table 1, SDHIs indicate drought magnitude (Figs. 2 and S3) or drought severity, the latter either expressed as volume-based anomaly or deficit (Figs. 3 and S4) or as frequency of non-exceedance (denoted with the suffix "_f") (Figs. 5 and S5). In addition, CQDI1(Q80) and CQDI1(Q80-HS) are compared at the global scale with respect to drought occurrence during the whole reference period (Fig. 4). SDHIs are further illustrated for four selected gauging stations with different streamflow regimes and assumed vulnerabilities of the risk system to streamflow anomalies (Figs. 6, S2 and S6). These include two stations with low interannual streamflow variability (Danube River at Hofkirchen, Germany (probably low vulnerability), and Angara River at Boguchany, Russia (possibly higher vulnerability)) and two stations with high interannual variability (White River near Oacoma, U.S. (probably low vulnerability), and Orange River at Vioolsdrif, South Africa (possibly higher vulnerability)). ~~SDHIs of drought magnitude (Figs. 2 and S3) and severity (Figs. 4) are compared separately (Sect. 4.2.1-4.2.3) following the classification system presented in Fig. 1. The SDHIs are shown in global maps for for a selected monthJuly 2003 and July 2015 with known drought events in Europe (INSERT REFERENCE). (March 2002), as it is important to understand the relation between indicators at a certain point in time, especially for the application in DEWS, which are focused on the current situation or the near future. As patterns of indicators depend on characteristic of the streamflow regime and water use that are temporally constant over the reference period, the reasons for similarities and differences between indicators can be deducted in any month of the reference period. March 2002 was selected as it was among the months with the highest difference between CQDI-Q80 and CQDI-Q80-HS. In addition to the analysis for the selected time step, CQDI1(Q80) and CQDI1-Q80-HS the latter two indica-~~

tors are compared at the global scale with respect to drought occurrence during the whole reference period (Fig. 4). Discrepancies and similarities between SDHIs are further illustrated using time series of SDHIs for selected grid cells with a certain streamflow regime and for inflow cells to reservoirs ((Rephrase): Each of the following sections is focussed on a core message leading to the selection of global-scale SDHIs in Sect. 5.

Discrepancies and similarities of the indicators are discussed in more detail for two gout illustrative grid cells with the same CQDI Q80 value in March 2002 (Sect. 4.2.4). reformulate objective: A comparison of SSI1 with SPI12 and SPEI12 is included in Sect. 4.1.x.... Finally, the suitability of SPI with different averaging periods to estimate streamflow drought hazard is assessed using streamflow observations from 218 GRDC gauging stations (Sect. 4.2.5).

Based on this global scale analysis and the proposed habituation based classification approach, selected SDHIs are recommended for implementation in large scale DEWS (Sect. 5).

4.1 Model validation

As a limited validation exercise in the present study, percent deviations of simulated Q80 from observed Q80 values were computed per calendar month (Fig. 3). The latter were based on monthly streamflow observations from the GRDC database (GRDC, 2019). Out of 1319 WaterGAP calibration stations, 220 stations with continuous monthly observations between 1986 and 2015 were assessed, which are mainly distributed over the Northern Hemisphere (U.S., Canada, Europe, Russia). Figure 9 depicts most of these stations. Two additional stations are located in the U.S. The analysis reveals that Q80 is overestimated by WaterGAP in 63% of all months and stations and in 53% if only relevant deviations > 10% are considered. The median percent deviation ranges between 35% in February and -7% in July. Figure 3 indicates a tendency of WaterGAP 2.2d to overestimate observed Q80 between October and April, while Q80 during the low-flow period in the Northern Hemisphere (May to September) is better captured.

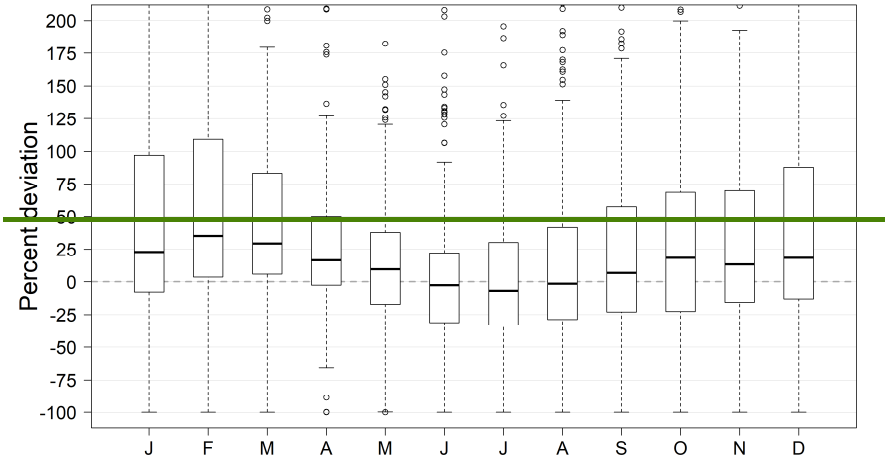


Figure 3: Percent deviations of simulated Q80 per calendar month from Q80 based on GRDC observations using the reference period 1986-2015.

~~In a recent study, WaterGAP 2.2d model output was validated against GRDC data by comparing SSI3 based on simulated and observed monthly streamflow (SSI3(sim) and SSI3(obs)) during 1971-2000 at 183 globally distributed GRDC stations (Wan et al., 2021). Applying drought hazard classes for SSI according to Agnew (2000), the agreement between simulated and observed hazard classes in each month was analyzed. Among all stations, the agreement ranged between 29 to 88% of all 360 months (their Fig. S4 and Table S3). At 68% of all stations (covering 83% of the assessed basin area), SSI3(sim) and SSI3(obs) resulted in the same drought hazard class in 70 to 88% of the time. Moreover, the goodness of fit was evaluated based on the Nash-Sutcliffe efficiency (NSE) for monthly streamflow and SSI3 (their Fig. S3). With a median NSE of 0.5 and an interquartile range of 0.2 0.7 for SSI3 and 0.14 0.7 for streamflow, WaterGAP 2.2d model output showed a moderate agreement with the observations. Both NSEs exceeded 0.7 at 25 out of the 183 stations, which are located in Central and Eastern Europe (twelve stations), the United States (ten stations), and South Africa (one station).~~

4.1 SDHIs based on empirical percentiles or standardized streamflow

EP1 patterns (Fig. 2c (July 2003), and Fig. S3c (July 2015)) are very similar to SSI1 (Fig. 2a and S3a) since both indicators are based on the same conceptual drought definition (Sect. 3.1). Both indicators generally identify the same drought regions. However, drought classes differ in many regions of the world, with EP1 indicating both higher and lower drought magnitude. For instance, in eastern France, EP1 indicates a higher drought magnitude class in July 2003 (return period RP > 20 years) than SSI1 (RP > 10 years), while it is vice versa in southern Germany. In July 2015, SSI1 indicates a higher drought hazard than EP1 in the Alpine region, but a lower hazard in parts of the Oder basin in eastern Poland. These differences can be attributed to the fitting of the gamma distribution in the case of SSI1 and the assignment of the maximum rank among tied values within a streamflow sample in the case of EP1 (Sect. 2.3~~2~~.43).

Comparing SSI1 with empirical percentiles, Tijdeman et al. (2020) identified several advantages and limitations for both indicators. SSI1 has the disadvantage that for different streamflow regimes, different parametric probability distributions would be required to achieve the best fit, which reduces consistency at the global scale. In this study, the gamma distribution showed the best fit among 23 parametric probability distributions for most grid cells and was applied in each month and grid cell. Of course, using only one distribution for the whole globe results in poorly fitting distributions for some cells and months (Tijdeman et al., 2020). Grid cells where gamma fitting was rejected in the calendar month July based on the KS test (Sect. 2.3.1) are shown in grey in Fig. 2a (18% of all grid cells excluding Greenland). EP1 does not require fitting of a distribution and can therefore be computed in more grid cells than SSI1. Only if a sample includes more zero flows than the selected threshold, drought identification is not possible (blue grid cells in Figs. 2c). On the other hand if Q80 is zero and the current streamflow exceeds zero, it is possible to define that the current month is not a drought month (shown in beige in Figs. 2a and 2c). EP1 has the disadvantage that it only allows the quantification of the historical non-exceedance frequency within the reference period, while probabilistic information, for example on extreme events such as a 100-year drought, cannot be derived (Tijdeman et al., 2020). Nonetheless, EP1 seems to be more suitable for a global-scale DEWS, as the indicator does not

entail the possibly large uncertainties due to the fitting of a probability distribution and can be computed in more grid cells than SSI1.

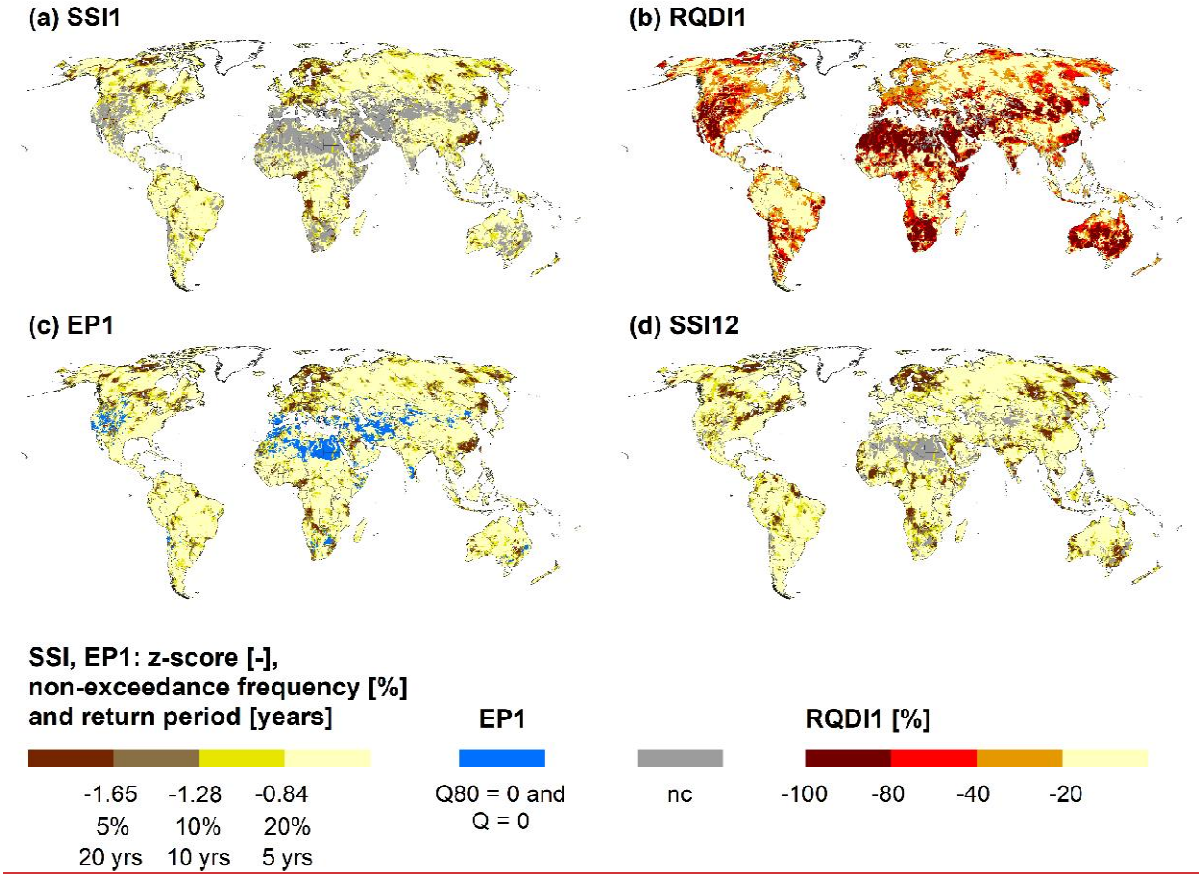


Figure 42: Magnitude of drought hazard (level 1 in Fig. 1): Non-cumulative anomaly in ~~March~~July 200~~2~~3 as indicated by SSI1 (a), RQDI1 (b), EP1 (c), and SSI12 (d)~~, SPI12 (e) and SPEI12 (f)~~ for the reference period 1986-2015. For the standardized indicators and EP1, the z scores and the corresponding frequencies of non-exceedance and return periods are shown. In the blue grid cells in (c), drought identification is not possible with EP1, since Q80 and Q are zero. "nc": not computable. ~~The two grid cells from Table 2 are marked in (a) (northern Italy and central Paraguay).~~

4.~~1~~2 SDHIs assuming habituation to mean streamflow or interannual variability of streamflow

With percentile-based indicators (e.g., EP1, SSI1), risk systems in different regions are assumed to be equally habituated to a certain interannual streamflow variability, which is most likely not the case as interannual variability varies strongly (Fig. A1b). Comparing two regions with high and low interannual variability, the same streamflow percentile or z-score corresponds to a much higher relative deviation from mean calendar month streamflow (RQDI1) if interannual variability is high.

For instance, at the Orange River and White River with high interannual variability (Fig. S2), SSI1 values below -0.84 (RP > 5 years) always correspond to RQDI1 values below -70% and -60%, respectively. At the Danube River and Angara River (low interannual variability), RQDI1 of -50% is (almost) never reached, while maximum SSI1 values are higher than at the Orange River and White River. Hence, SSI1 might underestimate drought magnitude if interannual variability is high, especially for vulnerable systems.

At the global scale, RQDI1 (Figs. 2b and S3b) identifies most of the drought hotspots as indicated by EP1 (Figs. 2c and S3c), although the relative levels of magnitude differ. These differences correspond well to the interannual streamflow variability depicted in Fig. A1b. Drought hotspots according to EP1 in regions with low interannual variability (parts of North America, northern Europe, northern Russia) only show moderate relative streamflow deviations by global comparison. This is because RQDI1 values of -50% or lower are never reached in these regions, as was illustrated at the Danube and Angara stations (see above). Here, RQDI1 might underestimate drought magnitude. On the other hand, in regions with high interannual variability (e.g., large parts of Africa, central Asia, western U.S.), both drought magnitude and the affected area are larger according to RQDI1. Here, RQDI1 can draw attention to potential drought impacts in regions with higher suspected vulnerability (e.g., southern Africa) that would otherwise be overlooked using EP1 or SSI1. In regions where people are probably well accustomed to the interannual variability of streamflow (e.g., western U.S.), RQDI1 is less suited than EP1 to indicate drought magnitude. At the severity level, regions with low interannual variability are excluded using CRQDI1(-50%)_f (Figs. 4d and S5d) due to the low threshold of -50% (grid cells in light grey).

~~Figure 4 compares indicators of drought magnitude for March 2002 with averaging periods of one month and twelve months. Comparing SSI1 and RQDI1 (Figs. 4a and b), the patterns are different in many parts of the globe. While RQDI1 identifies most of the drought regions according to SSI1 (e.g., western U.S. and Canada, parts of Brazil, Siberia, India, and China), relative levels of drought magnitude are very different. For instance, streamflow drought hazard in northern Siberia is extreme according to SSI1 (return period of 20 years or higher), but the deviation from mean monthly streamflow according to RQDI1 is comparably moderate (~20-40%). This is due to the low interannual variability of streamflow (Fig. A1b). Moreover, RQDI1 identifies severe drought hazards below -80% in many regions with high interannual streamflow variability that are not in drought according to SSI1 (e.g., southern Africa, Australia, and northeastern China). The strong correlation between RQDI1 and the interannual variability can be clearly seen by comparing Fig. 4b and Fig. A1b. Overall, RQDI1 values below -40% are computed for a rather high fraction of grid cells (33% excluding Greenland). Analyzing all March results during the reference period, this fraction varies between 29% and 40%. Figures 4a and 4b underline that RQDI1 can add value to global scale assessments by drawing the attention to highly vulnerable regions.~~

~~Application of CQDI1(Q50) (Fig. 5c) implies that the system at risk is only habituated to the seasonality of streamflow and that the study region is in drought half of the time. Like RQDI1 (Fig. 4b), the indicator can identify drought in highly vulnerable regions that would otherwise be overseen using lower thresholds such as Q80.At the global scale, CRQDI1(-50%)_f (Fig. 7c) identifies fewer regions with severe drought status compared to CQDI1(Q50)_f, but more regions compared to~~

CQDI1(Q80)_fFigure 7 depicts the probability (frequency) of non-exceedance p of drought severity in March 2002 between four CQDI1 variants, the cumulative relative deviation CRQDI1 with a threshold of -50%, and the cumulative empirical percentile CEP1 with a threshold of 20%. The indicators are denoted with the suffix "f" for frequency. A p value of 0.8, for example, indicates a high drought hazard, where the severity up to March 2002 is higher than the severity of 80% of all completed drought events in the reference period. Expressing severity in probability of non-exceedance, as also done in Cammalleri et al. (2016a), facilitates comparison between different indicator types that are quantified with different units.

4.2 SDHIs based on empirical percentiles or standardized streamflow

EP1 patterns (Fig. 4c) are very similar to SSI1, since both indicators are based on the same conceptual drought definition. Both indicators generally identify the same drought regions. However, drought classes differ for many grid cells with EP1 indicating both more and less severe droughts within each region. These differences are due to the fitting of the gamma distribution in case of SSI1 and due to the assignment of the maximum rank among tied values within a streamflow sample in case of EP1 (Sect. 2.2.4). Comparing SSI1 with empirical percentiles, Tijdeman et al. (2020) identify several advantages and limitations for both indicators. SSI1 has the disadvantage that for different streamflow regimes, different parametric probability distributions would be required to achieve the best fit, which reduces consistency at the global scale. In this study, the gamma distribution showed the best fit among 23 parametric probability distributions for most grid cells and was applied in each month and grid cell. Of course, using only one distribution for the whole globe results in poorly fitting distributions for some cells and months (Tijdeman et al., 2020) especially at the lower bound. Grid cells where gamma fitting was rejected in March based on the KS test (Sect. 2.2.2) are shown in grey in Fig. 4a (18% of all grid cells excluding Greenland). EP1 does not require fitting of a distribution and can therefore be computed in more grid cells than SSI1. Only if a sample includes more zero flows than the selected threshold, drought identification is not possible (blue grid cells in Figs. 4c). In Fig. 4a, these cells all coincide with grid cells where gamma fitting was rejected. On the other hand, if Q80 is zero and current streamflow exceeds zero, it is possible to define that the current month is not a drought month (shown in beige in Figs. 4a and 4c). EP1 has the disadvantage that it only allows the quantification of the historical non-exceedance frequency within the reference period, while probabilistic information, for example on extreme events such as a 100-year drought, cannot be derived (Tijdeman et al., 2020).

Although SSI1 and EP1 are based on the same conceptual drought concept, they indicate different drought anomalies. The streamflow volume in March 2002 is the fourth smallest value in the Italian cell (rank 4, EP1=4/30=0.13) and the second smallest value in the Paraguayan cell (rank 2, EP1=2/30=0.07). However, since the slope at the lower bound of the ranked streamflow values in the Italian cell (not shown) is very small compared to the latter cell, the non-exceedance probability p of the fitted gamma distribution increases equally slowly, and the resulting p (and z score) is smaller than in the Paraguayan cell. In fact, the smallest SSI1 in March in the Italian grid cell is -1.76 and in the Paraguayan cell -1.56. Hence, a severe drought, usually defined below SSI1= -1.65, is never identified in March in the latter grid cells.

~~Spatial patterns based on CQDI1(Q80)_f (Fig. 7a) and CEP1(20%)_f (Fig. 7f) are similar, but differences are visible in several regions. In southeastern Russia, northeastern China, Siberia, and parts of Canada and Alaska, CEP1(20%)_f indicates a more severe drought event than CQDI1(Q80)_f, while in the Mediterranean regions and the eastern and southwestern U.S. the severity of the drought events is lower according to CEP1(20%)_f. These differences occur since CQDI1(Q80) quantifies absolute and CEP1 relative drought deficits resulting in different relative levels of drought severity among the drought events (Sect. 4.2.4 and Fig. 8).~~

4.3 SDHIs taking into account human water use and EFR

The water deficit indicators ~~CQDI1_WUs~~CQDI1(WUs) and CQDI1(WUs-EFR) (Figs. ~~5~~3c,d and S4c,b) define drought as "less water than needed" as opposed to the anomaly indicator CQDI1(Q80) (Figs. 3a and S4a) indicating "less water than normal" (or rather less water in a certain month than in 80% of the years). Consequently, the spatial pattern of the former is very different from CQDI1(Q80) patterns. ~~shows a completely different spatial pattern from the above presented indicators, since it is driven by both the spatial pattern of water stress (human water demand for surface water as a fraction of mean streamflow, Fig. A1c) and low water availability.~~ For instance, the drought events in 2003 and 2015 in central and eastern Europe identified by CQDI1(Q80) are not indicated by CQDI1(WUs). This is because surface water stress is generally low in these regions. The spatial patterns of CQDI1(WUs) correlate well with Fig. A1c comparing human water demand for surface water as a fraction of mean streamflow. CQDI1(WUs-EFR) additionally considers the environmental flow requirement EFR computed as 80% of naturalized mean calendar month streamflow. Like RQDI1, the indicator thus depends on mean monthly streamflow and the spatial pattern corresponds well to the map of interannual variability (Fig. A1b). ~~while low water availability leads to high CQDI1_WUsCQDI1(WUs) values in northern South Africa and southeast Spain in March 2002, the cumulative streamflow anomaly according to CQDI Q80 is low in both regions.~~

A comparison between ~~CQDI1_WUs~~CQDI1(WUs) and ~~CQDI1_WUs_EFR~~CQDI1(WUs-EFR) ~~(Fig. 5d)~~ shows that only in a few regions human water demand is the dominant component determining the water deficit ~~in March 2002 (e.g., parts of North America, India, northeastern China, and Australia).~~ In most regions, EFR leads to high cumulative deficits even if seasonal human water demand is small (< 10% of available streamflow, Fig. A1c). ~~Since EFR depends on mean streamflow per calendar month, CQDI1_WUs_EFRCQDI1(WUs_EFR) shows very similar patterns to RQDI1 (Fig. 4b). CQDI1_WUs_EFR~~CQDI1(WUs-EFR) is the only indicator in this study that takes explicitly into account the health of the river ecosystem, an aspect that should be included in a global-scale DEWS. Alternatively, the cumulative anomaly-deficit indicator QDAI (Popat and Döll, 2021), considering EFR based on a similar approach, can inform decision-makers and water users about the drought hazard for water supply. In strongly altered flow regimes, where simulated anthropogenic monthly streamflow (Qant) is always below 80% of mean monthly naturalized streamflow (Qnat), time series of ~~CQDI1_WUs_EFR~~CQDI1(WUs-EFR) are continuously ~~increasing~~increasing, and it is not possible to distinguish drought events. In such cases, it is more meaningful to set EFR to 80% of mean monthly Qant, implying that the altered flow regime is the "new normal" (see also Table 1).

~~…the map [now Fig. 5b] focuses the viewer to grid cells with potential water deficits for human water supply (in particular irrigation)~~

~~CQDI1(Q50)_f (Fig. 7c) and CQDI1(WUs-EFR)_f (Fig. 7d) show high correspondence, as both imply similar assumptions about the habituation of the system at risk to the streamflow regime (see Table 1). Correspondence between these two indicators is higher than between CQDI1(Q50)_f and CQDI1(Q80)_f.~~

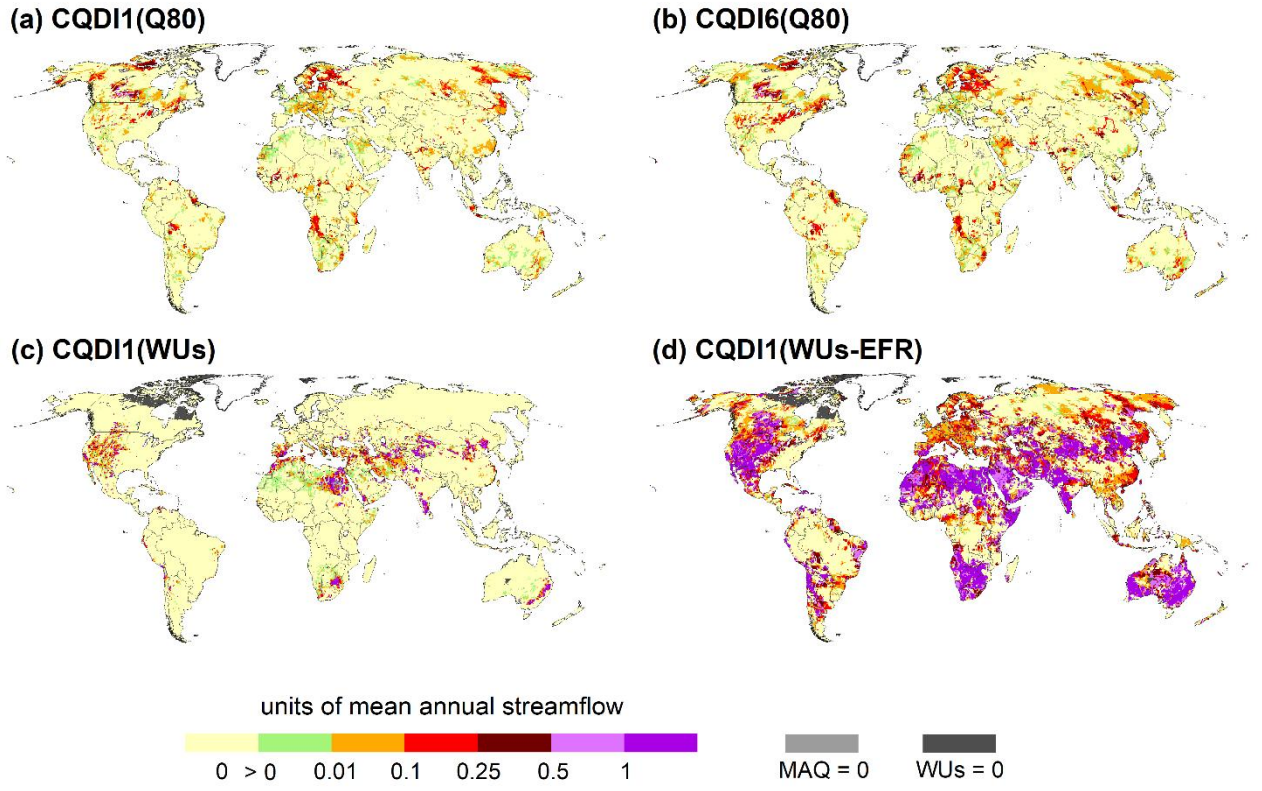
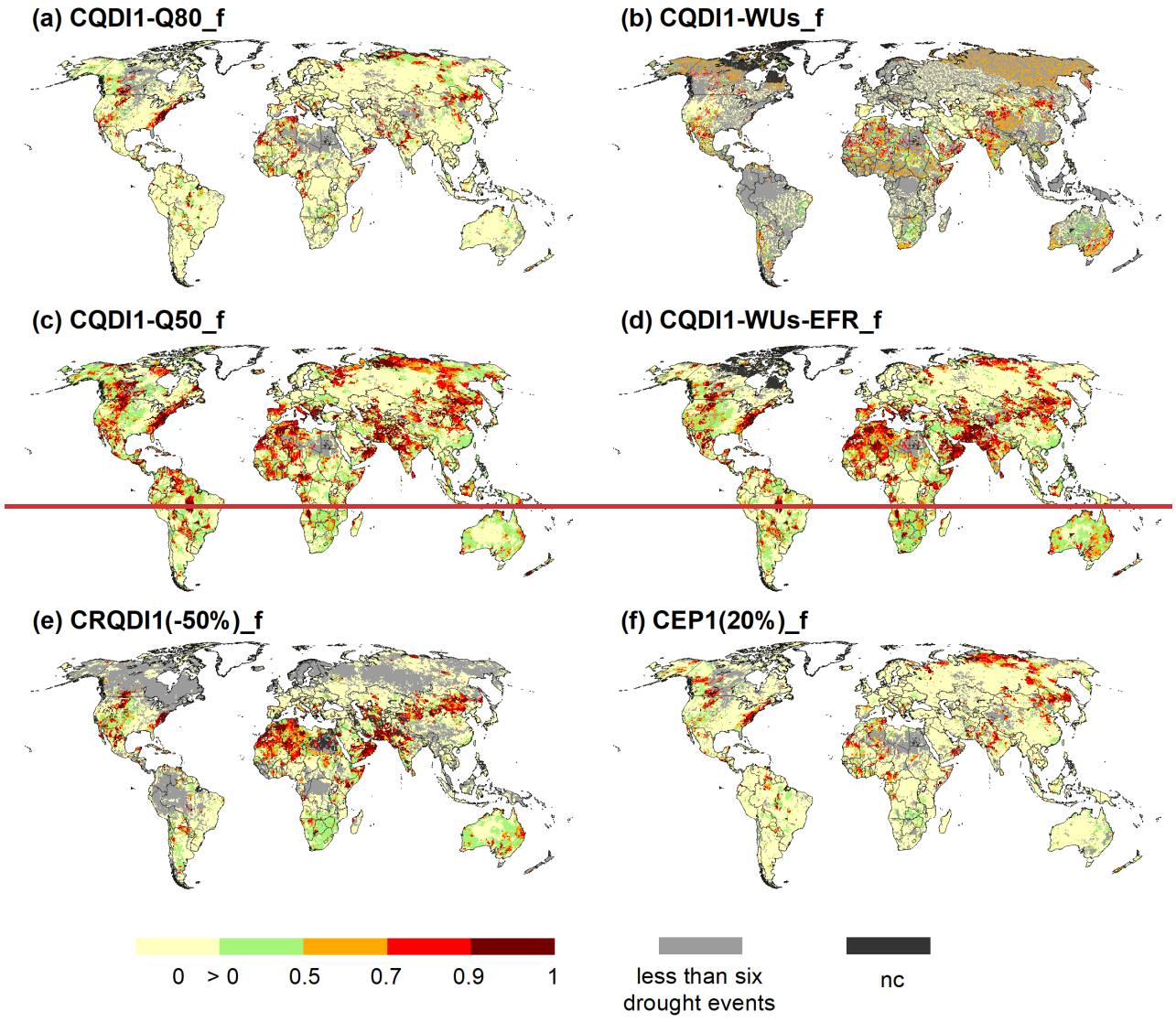


Figure 3: Severity of drought hazard (level 2 in Fig. 1): Cumulative deficit in July 2003 since onset of drought event as indicated by CQDI1(Q80) (a), CQDI6(Q80) (b), CQDI1(WUs) (c), and CQDI1(WUs-EFR) (d) for the reference period 1986-2015. Grid cells with a deficit of zero are shown in beige. Values larger than zero and below 0.1 are shown in green. A value of 0.1, for example, denotes that the current cumulative deficit is equivalent to 10% of mean annual streamflow (MAQ). WUs: mean annual surface water withdrawals.



4.4 SDHIs for reservoir management or water users with access to reservoirs

In large-scale hydrological modeling, it is very difficult to accurately simulate how man-made reservoirs affect water availability, i.e. how they impact downstream streamflow and how reservoir storage varies in time. For water users that depend on large reservoirs, streamflow deficits during the low-flow months are not relevant, since reservoirs can store water from the high-flow season. Hence, drought magnitude should be assessed using SDHIs with longer averaging periods that either assume habituation to interannual variability (e.g., SSI12, EP12, Table 1) or mean annual conditions (RQDI12, Table 1), but not seasonality. At the four investigated gauging stations (Fig. S2), the relation between SSI12 and RQDI12 is the same as for SSI1 and RQDI1 (Sect. 4.2). If interannual variability is high (Orange River and White River), SSI12 values correspond to

much higher RQDI12 values compared to the stations with low interannual variability (Danube River and Angara River). To obtain drought severity, these indicators can be cumulated using a suitable threshold. As described in Sect. 4.2, a threshold of -50% for RQDI12 would exclude regions with low interannual variability, where this value is rarely reached, and where RQDI12 might underestimate drought magnitude.

In addition to CSSI12 (or CEP12) and CRQDI12, the volume-based indicators CQDI1(Q80-HS) and CQDI6(Q80) were assessed.

With CQDI1(Q80-HS), an existing drought is allowed to continue in months where the calendar month Q80 is zero, even if streamflow Q exceeds zero. In contrast, CQDI1(Q80) only allows a drought to continue if Q80 and Q are zero.

A comparison of ~~CQDI1_Q80~~CQDI1(Q80) and ~~CQDI1(Q80-HS)~~the two indicators ~~CQDI1(Q80-HS)~~ (Fig. ~~6~~4) reveals that the impact of the HS method ~~(Sect. 2.2.3)~~ is rather small at the global scale but can be relevant at the regional scale. Figure ~~6~~4a depicts the fraction of drought months as a percentage of all 360 months during the reference period as indicated by ~~CQDI1_Q80~~CQDI1(Q80). Using Q80 as threshold implies that the time series should be in drought 20% of the time. ~~This is only the case in 6% of all grid cells (excluding Greenland), while in 86% the fraction is reduced to the range of >0% to <20%. This is either due to the fact that one month droughts are ignored (see 2mc, Sect. 2.2.3) or that several calendar months with Q80=0 exist where a streamflow deficit can never be identified. The fraction is increased to up to 22% in 5% of all grid cells either due to the pooling of drought events (2mc) or due to drought prolongation in case of Q=0 and Q80=0 (Sect. 2.2.3).~~The fact that this percentage is often reduced and sometimes increased can be attributed to the two months criterion (Sect. 2.3.2) (one month droughts are ignored and several droughts are pooled) and to drought prolongation if Q80 and Q are zero. ~~Furthermore, CQDI1_Q80~~CQDI1(Q80) ~~is always zero in 3% of all grid cells, where Q80 is zero in many calendar months and one month droughts are ignored in the remaining months. In conclusion, due to the assumed habituation of people and the ecosystem to periods of zero streamflow and to very short streamflow deficits, streamflow drought hazard as quantified by CQDI1_Q80~~CQDI1(Q80) ~~is less frequent in the grey, green, and beige grid cells (Fig. 6a). Regions where drought occurrence is reduced to less than 14% of the time include Japan, large parts of China, Pakistan, Afghanistan, Iran, North Africa, the western parts of South and North America, and eastern Australia.~~ The HS method leads to an increase in drought months by up to 3 percent points (corresponding to 11 out of 360 months) in 6% of all grid cells, ~~e~~e.g., parts of India, Pakistan, Afghanistan, Iran, and the western U.S., all of which are regions with highly seasonal streamflow regimes (Fig. ~~6~~4b). Larger increases of up to 12 percent points are only computed in 0.4% of all grid cells~~. Higher values for the CQDI1(Q80-HS)~~CQDI1(Q80-HS) ~~in~~. Hence, the additional information value of CQDI1(Q80-HS) in a large-scale DEWS would be small. Instead, CQDI variants with longer averaging periods like CQDI6(Q80) (Figs. 3b and S4b) are more suitable for assessing risk systems with reservoirs. The time series of CQDI6(Q80) at the four gauging stations (Fig. S2) illustrate how the maximum drought severity is shifted by one month or more compared to CQDI1(Q80), reflecting that a reservoir storage requires several months of "normal" streamflow to be replenished.

~~dicator are computed in parts of India, Pakistan, Afghanistan, Iran, and the western U.S., all of which are regions with highly seasonal streamflow regimes (Fig. A1a). Hence, although the HS method has a small effect at the global scale, differ-~~

~~ences between CQDI1-Q80~~CQDI1(Q80) and ~~CQDI1(Q80-HS)~~CQDI1(Q80-HS) ~~can be significant in regions with high seasonal streamflow variability.~~

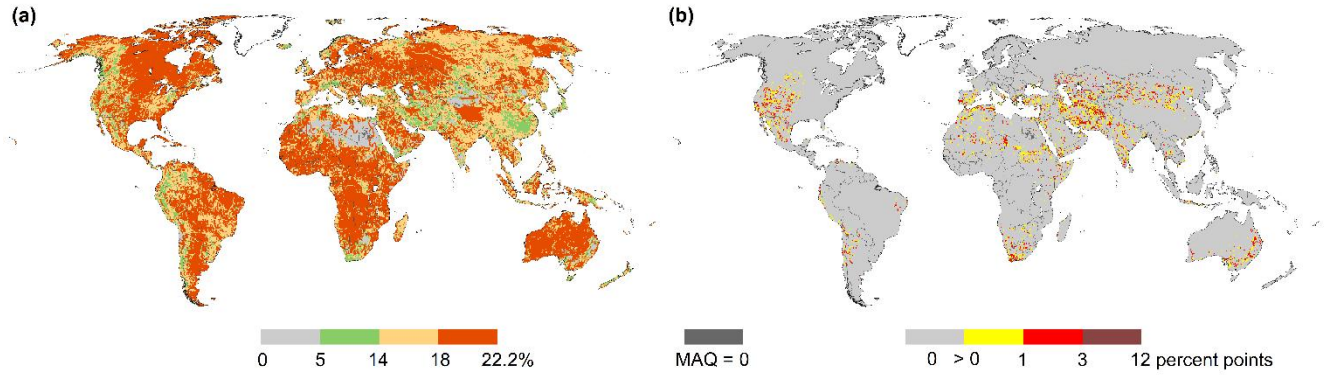


Figure 4: Comparison of CQDI1(Q80) and CQDI1(Q80-HS) in the reference period 1986-2015: Percent of months in drought based on CQDI1(Q80) (a) and the increase due to the "HS method" in percent points (b). Both indicators allow an existing drought to continue in months where Q80 and the current streamflow Q are zero. The HS method additionally facilitates drought prolongation in months with Q80=0 if Q>0. Neither indicator allows a drought to begin in months with Q80=0. Drought prolongation in the case of Q80=0 is only possible if a streamflow deficit was computed in at least two antecedent months with Q80>0 (2mc, Sect. 2.~~2~~3.~~3~~2). In (a), the fraction of drought months is reduced to <20% if one-month droughts are ignored (2mc). In grid cells with 0% in (a), Q80 is either always zero, or the few calendar months with Q80>0 result in one-month droughts only. The fraction can be increased to >20% in case of drought pooling (2mc) or in the case of drought prolongation if Q80=0. MAQ: mean annual streamflow.

~~SSI12 (Fig. 4d) indicates where average streamflow between April 2001 and March 2002 is very low compared to April to March periods during 1986-2015. Compared to SSI1, the areal extent of extreme drought magnitude (< -1.65) as identified by SSI12 is larger in, e.g., central Brazil, Morocco, north eastern China, Siberia, and Greece. If people in these regions need streamflow to fill reservoirs, SSI12 is more suitable than SSI1 to detect the drought hazard. In other parts of the globe, the areal extent of extreme drought magnitude is smaller according to SSI12 (e.g., North America, northern Italy, and southern Africa). Here, SSI1 might detect the onset of a streamflow drought that cannot be captured by SSI12. At the global scale, it is unknown if people depend directly on streamflow or if they have access to reservoirs. Therefore, a global scale DEWS should provide hazard indicators for both risk systems.~~

~~**4.6 SDHIs based on meteorological variables**~~

~~SPI12 and SPEI12 (Figs. 4e and f) can be used as proxies for identifying streamflow drought hazard if streamflow data is not available.~~In ~~a limited sensitivity analysis, the correlation between the observation based SSI1 (SSI1obs) and the five indicators SPI3, 6, 9, 12 and simulated SSI1 (SSI1sim) was computed at 220 gauging stations with continuous monthly~~

streamflow during the reference period (Supplementary material S4). Among the 217 stations where fitting of the gamma function was possible, the Pearson correlation coefficient r was highest for SSI1(sim) at 147 stations (68%) with a median r of 0.75, followed by SPI12 at 31 stations (median r = 0.59). In Fig. S7, the drought hazard indicator that achieved the highest r is indicated for the 217 gauging stations. Hence, drought hazard indicators based on modeled streamflow can often better quantify observed streamflow drought hazard than meteorological indicators. However, this strongly depends on the goodness of fit after streamflow calibration.

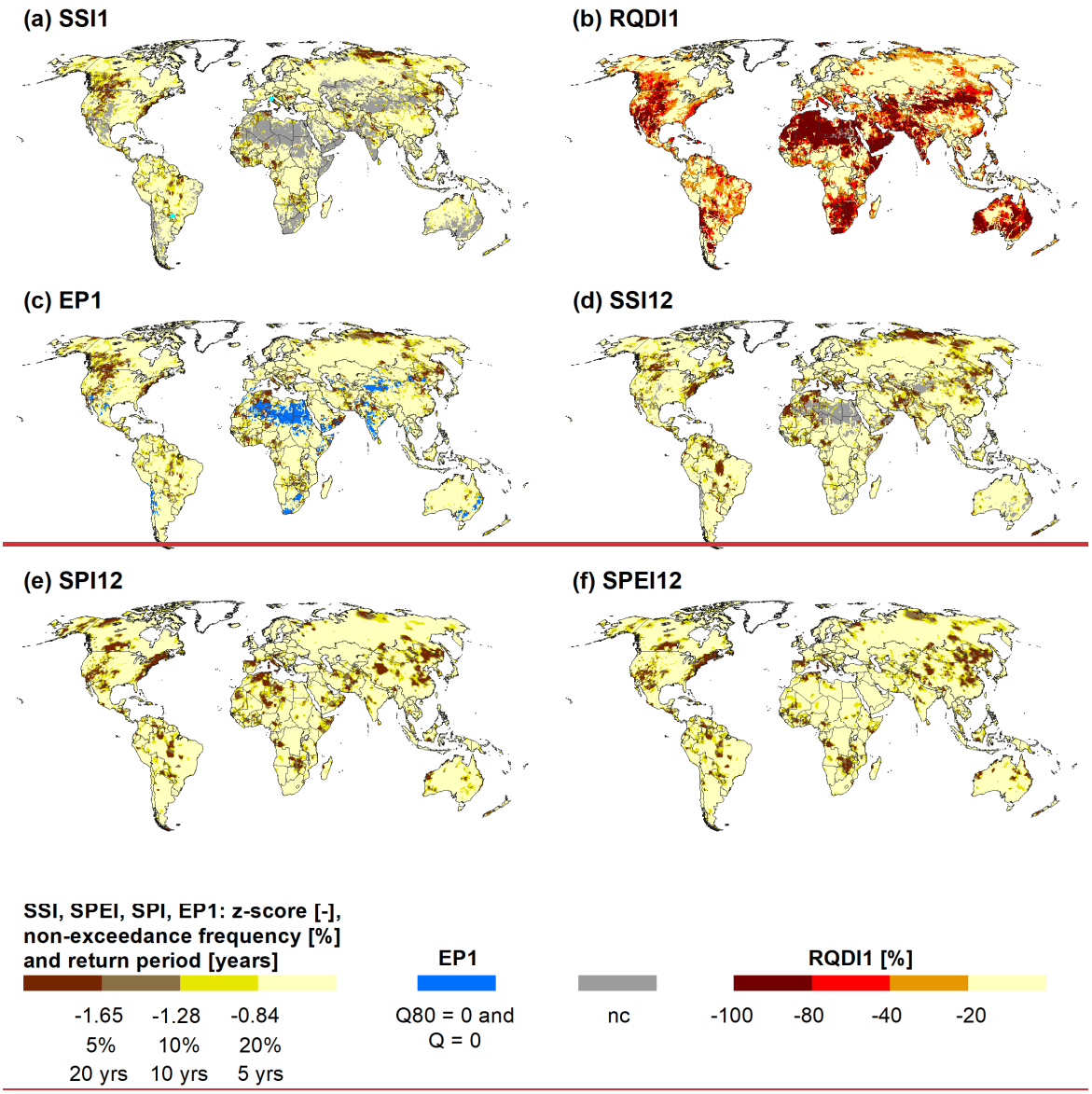
Focusing on the SPI variants only, the correlation with SSI1(obs) was highest for SPI12 and SPI3 in 30% and 26% of all basins followed by SPI6 and SPI9 (both ca. 22% of all basins). The SPI averaging period was found to be positively correlated with the basin size. In conclusion, SPI and SPEI should be used as proxies for identifying streamflow drought hazard in regions where the performance of WaterGAP is often lower. These include semi-arid basins, in particular where streamflow is highly altered by irrigation and man-made reservoirs, and in regions dominated by lakes. Of course, the correlation with SSI1 strongly depends on the selected averaging period, which varies with different basin characteristics (Sect. 4.2.5). Here, the selected indicators correlate fairly well with SSI1 by visual inspection. However, the areal extent of extreme drought magnitude below -1.65 is higher according to the proxy indicators (e.g., U.S. east coast, southern Africa, and eastern China). Hence, in a global assessment, different averaging periods for SPI and SPEI should be provided either at the global scale or specific to basins based on a correlation analysis.

4.2.1 Drought magnitude (level 1)

EP1 patterns (Fig. 4c) are very similar to SSI1, since both indicators are based on the same conceptual drought definition. Both indicators generally identify the same drought regions. However, drought classes differ for many grid cells with EP1 indicating both more and less severe droughts within each region. These differences are due to the fitting of the gamma distribution in case of SSI1 and due to the assignment of the maximum rank among tied values within a streamflow sample in case of EP1 (Sect. 2.2.4). Comparing SSI1 with empirical percentiles, Tijdeman et al. (2020) identify several advantages and limitations for both indicators. SSI1 has the disadvantage that for different streamflow regimes, different parametric probability distributions would be required to achieve the best fit, which reduces consistency at the global scale. In this study, the gamma distribution showed the best fit among 23 parametric probability distributions for most grid cells and was applied in each month and grid cell. Of course, using only one distribution for the whole globe results in poorly fitting distributions for some cells and months (Tijdeman et al., 2020) especially at the lower bound. Grid cells where gamma fitting was rejected in March based on the KS test (Sect. 2.2.2) are shown in grey in Fig. 4a (18% of all grid cells excluding Greenland). EP1 does not require fitting of a distribution and can therefore be computed in more grid cells than SSI1. Only if a sample includes more zero flows than the selected threshold, drought identification is not possible (blue grid cells in Figs. 4c). In Fig. 4a, these cells all coincide with grid cells where gamma fitting was rejected. On the other hand, if Q80 is zero and current streamflow exceeds zero, it is possible to define that the current month is not a drought month (shown in beige in Figs. 4a and 4c). EP1

has the disadvantage that it only allows the quantification of the historical non-exceedance frequency within the reference period, while probabilistic information, for example on extreme events such as a 100-year drought, cannot be derived (Tijdeman et al., 2020).

SPI12 and SPEI12 (Figs. 4e and f) can be used as proxies for identifying streamflow drought hazard if streamflow data is not available. Of course, the correlation with SSI1 strongly depends on the selected averaging period, which varies with different basin characteristics (Sect. 4.2.5). Here, the selected indicators correlate fairly well with SSI1 by visual inspection. However, the areal extent of extreme drought magnitude below -1.65 is higher according to the proxy indicators (e.g., U.S. east coast, southern Africa, and eastern China). Hence, in a global assessment, different averaging periods for SPI and SPEI should be provided either at the global scale or specific to basins based on a correlation analysis. SSI12 (Fig. 4d) indicates where average streamflow between April 2001 and March 2002 is very low compared to April-to-March periods during 1986-2015. Compared to SSI1, the areal extent of extreme drought magnitude ($<$ -1.65) as identified by SSI12 is larger in, e.g., central Brazil, Morocco, north-eastern China, Siberia, and Greece. If people in these regions need streamflow to fill reservoirs, SSI12 is more suitable than SSI1 to detect the drought hazard. In other parts of the globe, the areal extent of extreme drought magnitude is smaller according to SSI12 (e.g., North America, northern Italy, and southern Africa). Here, SSI1 might detect the onset of a streamflow drought that cannot be captured by SSI12. At the global scale, it is unknown if people depend directly on streamflow or if they have access to reservoirs. Therefore, a global-scale DEWS should provide hazard indicators for both risk systems. For monitoring drought risk for river ecosystems, short averaging periods are more suitable assuming that the ecosystem is habituated to seasonality.



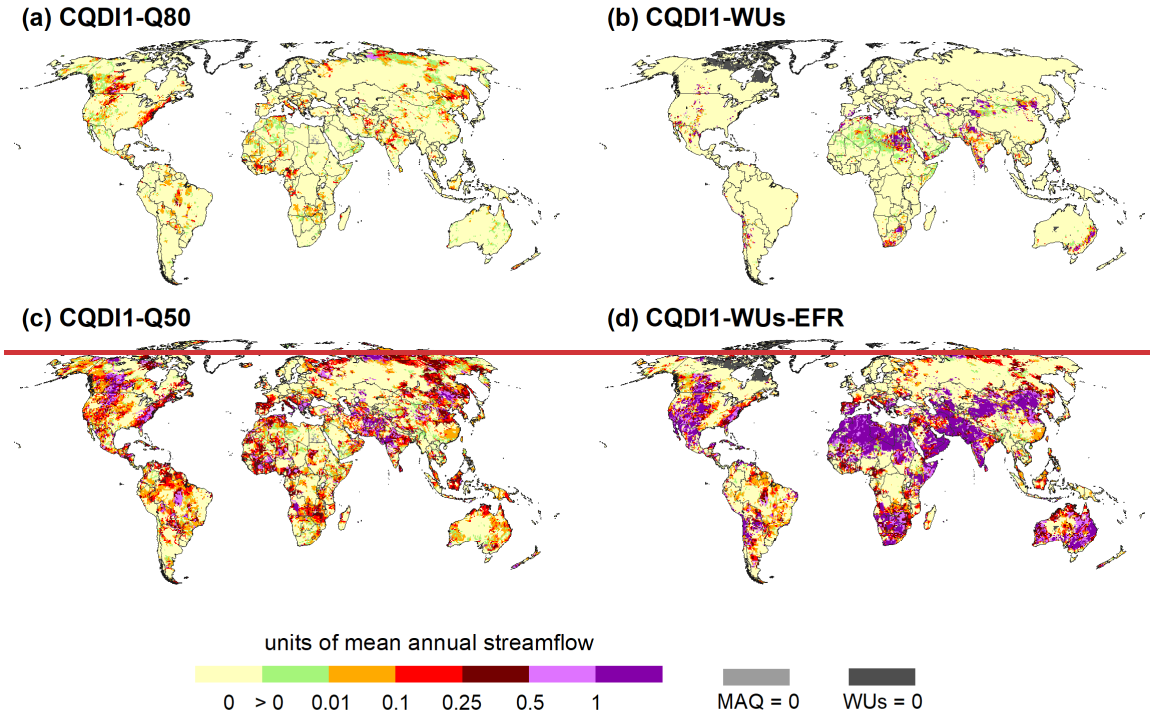
~~**Figure 4: Magnitude of drought hazard (level 1 in Fig. 1): Non-cumulative anomaly in March 2002 as indicated by SSI1 (a), RQDI1 (b), EP1 (c), SSI12 (d), SPI12 (e) and SPEI12 (f) for the reference period 1986-2015. For the standardized indicators and EP1, the z scores and the corresponding frequencies of non-exceedance and return periods are shown. In the blue grid cells in (c), drought identification is not possible with EP1, since Q80 and Q are zero. "nc": not computable. The two grid cells from Table 2 are marked in (a) (northern Italy and central Paraguay).**~~

~~4.2.2 Drought severity (level 2)~~

~~Figure 5 shows volume-based indicators of drought severity in March 2002. Since CQDI1-Q80 (Fig. 5a) is a percentile-based indicator such as SSI1 and EP1 (Figs. 4a and 4c), the spatial patterns are similar. Nevertheless, as the severity indicator in-~~

cludes information on drought development before March 2002, while the magnitude indicators only quantify the drought condition in this month, the severity shows a more differentiated picture of drought conditions in areas with similarly strong drought according to SSI1 and EP1, e.g., Western North America and Northern Siberia. Therefore, drought anomaly indicators and drought severity indicators, either the volume-based version such a CQDI-Q80 or one based on EP1 (CEP1 in Fig. 1), provide different drought hazard information, and can therefore be used complementarily in a DEWS.

A comparison of CQDI1-Q80 and CQDI1(Q80_HS) (Fig. 6) reveals that the impact of the HS-method (Sect. 2.2.3) is rather small at the global scale but can be relevant at the regional scale. Figure 6a depicts the fraction of drought months as a percentage of all 360 months during the reference period as indicated by CQDI1-Q80. Using Q80 as threshold implies that the time series should be in drought 20% of the time. This is only the case in 6% of all grid cells (excluding Greenland), while in 86% the fraction is reduced to the range of >0% to <20%. This is either due to the fact that one month droughts are ignored (see 2mc, Sect. 2.2.3) or that several calendar months with Q80=0 exist where a streamflow deficit can never be identified. The fraction is increased to up to 22% in 5% of all grid cells either due to the pooling of drought events (2mc) or due to drought prolongation in case of Q=0 and Q80=0 (Sect. 2.2.3). Furthermore, CQDI1-Q80 is always zero in 3% of all grid cells, where Q80 is zero in many calendar months and one month droughts are ignored in the remaining months. In conclusion, due to the assumed habituation of people and the ecosystem to periods of zero streamflow and to very short streamflow deficits, streamflow drought hazard as quantified by CQDI1-Q80 is less frequent in the grey, green, and beige grid cells (Fig. 6a). Regions where drought occurrence is reduced to less than 14% of the time include Japan, large parts of China, Pakistan, Afghanistan, Iran, North Africa, the western parts of South and North America, and eastern Australia. The HS-method leads to an increase in drought months by up to 3 percent points (corresponding to 11 out of 360 months) in 6% of all grid cells (Fig. 6b). Larger increases of up to 12 percent points are only computed in 0.4% of all grid cells. Higher values for the CQDI1(Q80_HS) indicator are computed in parts of India, Pakistan, Afghanistan, Iran, and the western U.S., all of which are regions with highly seasonal streamflow regimes (Fig. A1a). Hence, although the HS-method has a small effect at the global scale, differences between CQDI1-Q80 and CQDI1(Q80_HS) can be significant in regions with high seasonal streamflow variability.



~~Figure 5: Severity of drought hazard (level 2 in Fig. 1): Cumulative deficit in March 2002 since onset of drought event as indicated by CQDI1-Q80~~CQDI1(Q80) (a), ~~CQDI1-WUs~~CQDI1(WUs) (b), ~~CQDI1-Q50~~CQDI1(Q50) (c), and ~~CQDI1-WUs-EFR~~CQDI1(WUs-EFR) (d) for the reference period 1986-2015. Grid cells with a deficit of zero are shown in beige. Values larger than zero and below 0.1 are shown in green. A value of 0.1, for example, denotes that the current cumulative deficit is equivalent to 10% of mean annual streamflow (MAQ). WUs: mean annual surface water withdrawals.

~~Application of CQDI1-Q50 (Fig. 5c) implies that the system at risk is only habituated to the seasonality of streamflow and that the study region is in drought half of the time. Like RQDI1 (Fig. 4b), the indicator can identify drought in highly vulnerable regions that would otherwise be overseen using lower thresholds such as Q80. The water deficit indicator CQDI1-WUs (Fig. 5b) shows a completely different spatial pattern from the above presented indicators, since it is driven by both the spatial pattern of water stress (human water demand for surface water as a fraction of mean streamflow, Fig. A1c) and low water availability. For instance, while low water availability leads to high CQDI1-WUs values in northern South Africa and south-east Spain in March 2002, the cumulative streamflow anomaly according to CQDI-Q80 is low in both regions.~~

~~A comparison between CQDI1-WUs and CQDI1-WUs-EFR (Fig. 5d) shows that only in a few regions human water demand is the dominant component determining the water deficit in March 2002 (e.g., parts of North America, India, northeastern China, and Australia). In most regions, EFR leads to high cumulative deficits even if seasonal human water demand is small (< 10% of available streamflow, Fig. A1c). Since EFR depends on mean streamflow per calendar month, CQDI1-WUs-EFR shows very similar patterns to RQDI1 (Fig. 4b). CQDI1-WUs-EFR is the only indicator in this study that takes into account~~

ecosystem health, an aspect that should be included in a global-scale DEWS. Alternatively, the cumulative anomaly-deficit indicator QDAI (Popat and Döll, 2021), considering EFR based on a similar approach, can inform decision-makers and water users about the drought hazard for water supply. In strongly altered flow regimes, where simulated anthropogenic monthly streamflow (Qant) is always below 80% of mean monthly naturalized streamflow (Qnat), time series of CQDI1-WUs-EFR are continuously increasing and it is not possible to distinguish drought events. In such cases, it is more meaningful to set EFR to 80% of mean monthly Qant implying that the altered flow regime is the "new normal status" (see also Table 1).

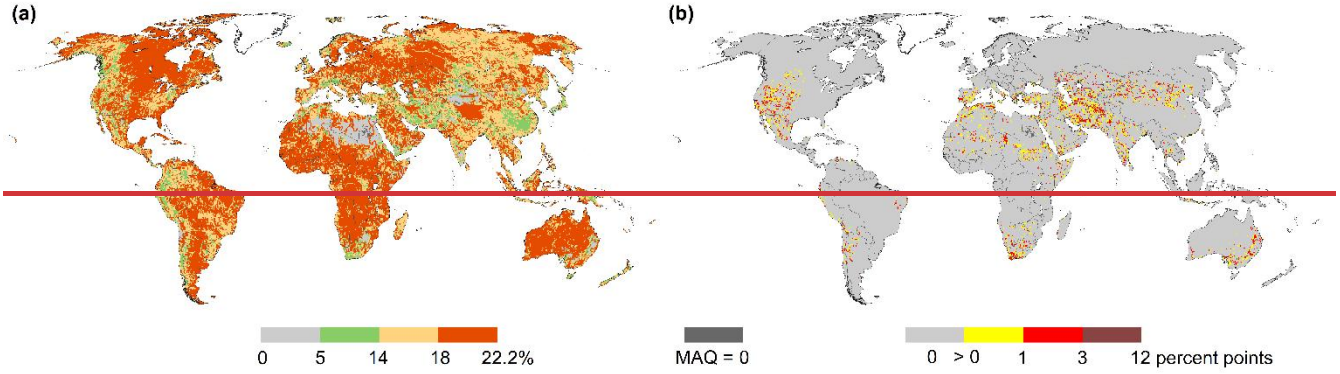
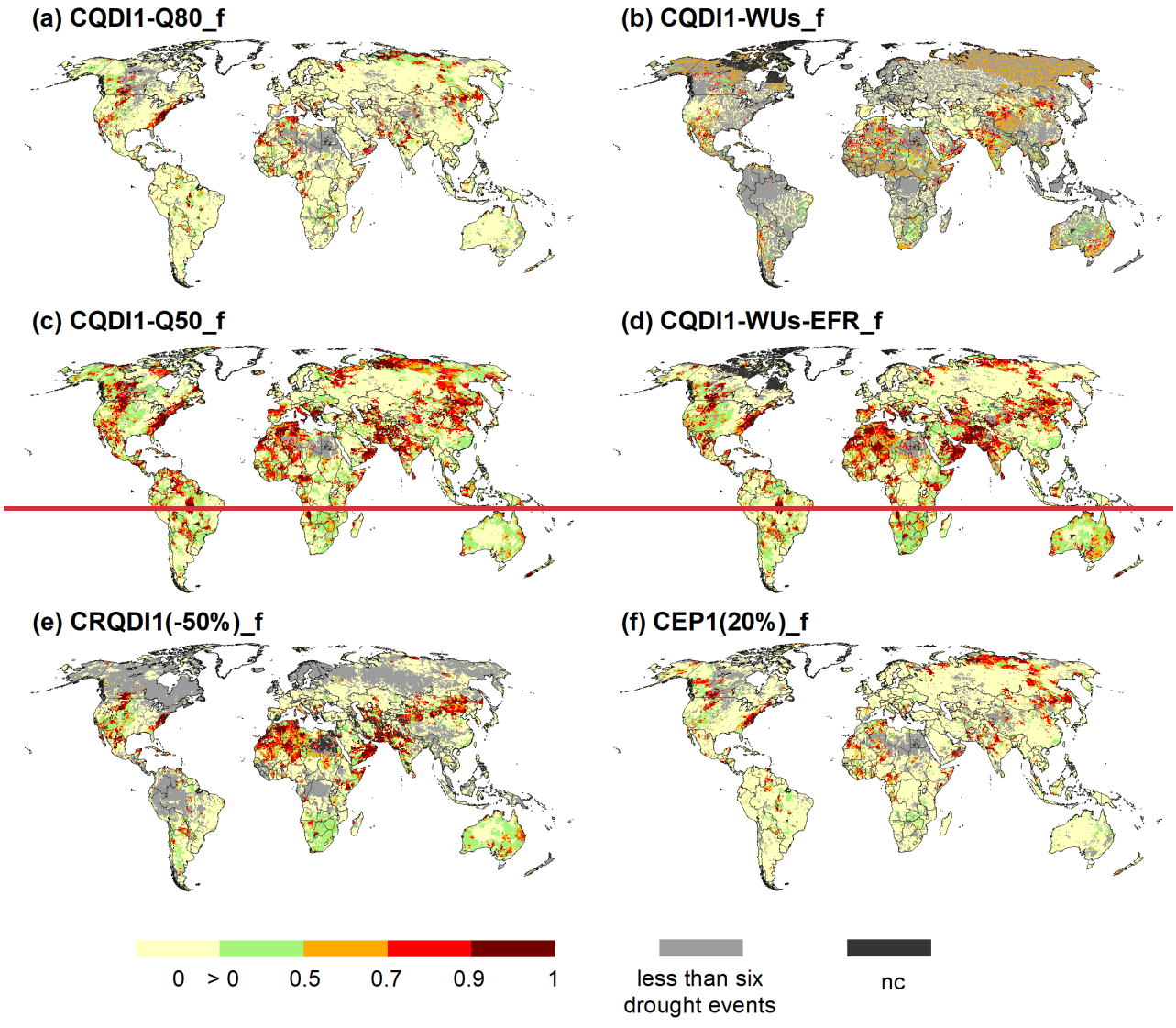


Figure 6: Comparison of CQDI1-Q80CQDI1(Q80) and CQDI1(Q80-HS)CQDI1(Q80-HS) in the reference period 1986-2015: Percent of months in drought based on CQDI1-Q80CQDI1(Q80) (a) and the increase due to the "HS-method" in percent points (b). Both indicators allow an existing drought to continue in months where Q80 and the current streamflow Q are zero. The HS method additionally facilitates drought prolongation in months with Q80=0 if Q>0. Neither indicator allows a drought to begin in months with Q80=0. Drought prolongation in case of Q80=0 is only possible if a streamflow deficit was computed in at least two antecedent months with Q80>0 (2mc, Sect. 2.2.3). In (a), the fraction of drought months is reduced to <20% if one month droughts are ignored (2mc). In grid cells with 0% in (a), Q80 is either always zero or the few calendar months with Q80>0 result in one month droughts only. The fraction can be increased to >20% in case of drought pooling (2mc) or in case of drought prolongation if Q80=0. MAQ: mean annual streamflow.

4.2.3 Drought severity expressed as frequency of non-exceedance (level 2)

Figure 7 depicts the probability (frequency) of non-exceedance p of drought severity in March 2002 between four CQDI1 variants, the cumulative relative deviation CRQDI1 with a threshold of -50%, and the cumulative empirical percentile CEP1 with a threshold of 20%. The indicators are denoted with the suffix "f" for frequency. A p value of 0.8, for example, indicates a high drought hazard, where the severity up to March 2002 is higher than the severity of 80% of all completed drought events in the reference period. Expressing severity in probability of non-exceedance, as also done in Cammalleri et al. (2016a), facilitates comparison between different indicator types that are quantified with different units. Spatial patterns based on CQDI1-Q80_f (Fig. 7a) and CEP1(20%)_f (Fig. 7f) are similar, but differences are visible in several regions. In southeastern Russia, northeastern China, Siberia, and parts of Canada and Alaska, CEP1(20%)_f indicates a more severe drought event than CQDI1-Q80_f, while in the Mediterranean regions and the eastern and southwestern U.S. the severity of the drought

events is lower according to CEP1(20%)_f. These differences occur since CQDI1-Q80 quantifies absolute and CEP1 relative drought deficits resulting in different relative levels of drought severity among the drought events (Sect. 4.2.4 and Fig. 8). For CQDI1_WUs_f (Fig. 7b), p values could not be computed in almost half of the grid cells, where less than six drought events were identified such that the map focuses the viewer to grid cells with potential water deficits for human water supply (in particular irrigation). CQDI1_Q50_f (Fig. 7c) and CQDI1_WUs_EFR_f (Fig. 7d) show high correspondence, as both imply similar assumptions about the habituation of the system at risk to the streamflow regime (see Table 1). Correspondence between these two indicators is higher than between CQDI1_Q50_f and CQDI1_Q80_f. At the global scale, CRQDI1(-50%)_f (Fig. 7e) identifies fewer regions with severe drought status compared to CQDI1_Q50_f, but more regions compared to CQDI1_Q80_f.



~~Figure 7: Probability of non-exceedance of drought events (level 2 in Fig. 1) in March 2002 for the cumulative indicators CQDI1-Q80CQDI1(Q80)_f (a), CQDI1-WUsCQDI1(WUs)_f (b), CQDI1-Q50CQDI1(Q50)_f (c), CQDI1-WUs-EFRCQDI1(WUs-EFR)_f (d), CRQDI1(-50%)_f (e), and CEP1(20%)_f (f) for the reference period 1986-2015. A value of 0.8, for example, indicates that the cumulative anomaly or deficit, i.e., the severity up to this month, is higher than the severity of 80% of all drought events in the reference period. The probability of non-exceedance was not computed for grid cells shown in light grey, where less than six drought events were computed in the reference period (Sect. 2.3). "nc": not computable.~~

4.~~752.4~~ Range of drought ~~hazard~~severity as quantified by ~~Relation between~~ the various SDHIs

A direct comparison between different severity indicators is possible when the time series of drought severity are transformed into frequency of non-exceedance.

Figures 5 and S5 depict the probability (frequency) of non-exceedance p of drought severity in July 2003 and July 2015, respectively, between four CQDI1 variants, the cumulative relative deviation CRQDI1 with a threshold of -50%, and the cumulative empirical percentile CEP1 with a threshold of 20%. The indicators are denoted with the suffix "f" for frequency. A p-value of 0.7, for example, indicates a high drought hazard, where the severity up to July 2003 is higher than the severity of 70% of all completed drought events in the reference period. In both example months, the spatial extent of regions with $p > 0.7$, i.e. severe droughts, is larger according to the indicators that do not assume habituation to interannual variability (CQDI1(Q50)_f, CQDI1(WUs)_EFR_f, and CRQDI1(-50%)_f). CQDI1(Q50)_f and CQDI1(WUs-EFR)_f are rather similar. Correspondence between these two indicators is higher than between CQDI1(Q50)_f and CQDI1(Q80)_f. CRQDI1(-50%)_f identifies fewer regions with severe drought status compared to CQDI1(Q50)_f, but more regions compared to CQDI1(Q80)_f.

Spatial patterns of CQDI1(Q80)_f (Figs. 5a and S5a) and CEP1(20%)_f (Figs. 5b and S5b) are very similar, since they are based on the same drought concept. Nonetheless, small differences occur in all identified drought hotspots, which can be explained by the fact that the former quantifies absolute and the latter relative streamflow anomalies per calendar month leading to a different ranking of low-flow and high-flow droughts during the reference period. This relation is illustrated for the Danube gauging station in Fig. 6 and for the other three investigated stations in Figs. S6~~a-c~~. Although CEP1(20%) (in units of cumulative percent) and CQDI1(Q80) (in units of mean annual streamflow) capture the same drought signal at the four stations, the relative levels among the drought events differ. In Fig. 6, the three most severe droughts according to CQDI1(Q80) are the drought events in 1998, followed by 2014 and 2003. In contrast, the 2003 drought, which occurred mainly during the low-flow period (August to November) has the second highest severity according to CEP1(20%). The high-flow drought from March to May 2011, on the other hand, has a lower severity rank according to CEP1(20%). The differences are more pronounced with higher seasonal variability (Orange River and White River, Figs. S6~~b and c~~), but almost negligible if seasonality is very low (Angara River, Fig. S6~~a~~). Consequently, in a large-scale DEWS, CEP1(20%) appears to be more suitable in re-

gions where the risk system is more vulnerable to low-flow droughts than to high-flow droughts. These differences would not occur if volume-based monthly streamflow deficits would be normalized using mean monthly streamflow but do occur if they are either not normalized (e.g., the low-flow index LFI, Cammalleri et al., 2016a) or normalized against mean annual streamflow volume (e.g., van Loon et al. (2014) and all CQDI1 variants in this paper).

~~At the global scale, CRQDI1(-50%)_f (Fig. 7e) identifies fewer regions with severe drought status compared to CQDI1(Q50)_f, but more regions compared to CQDI1(Q80)_f.~~

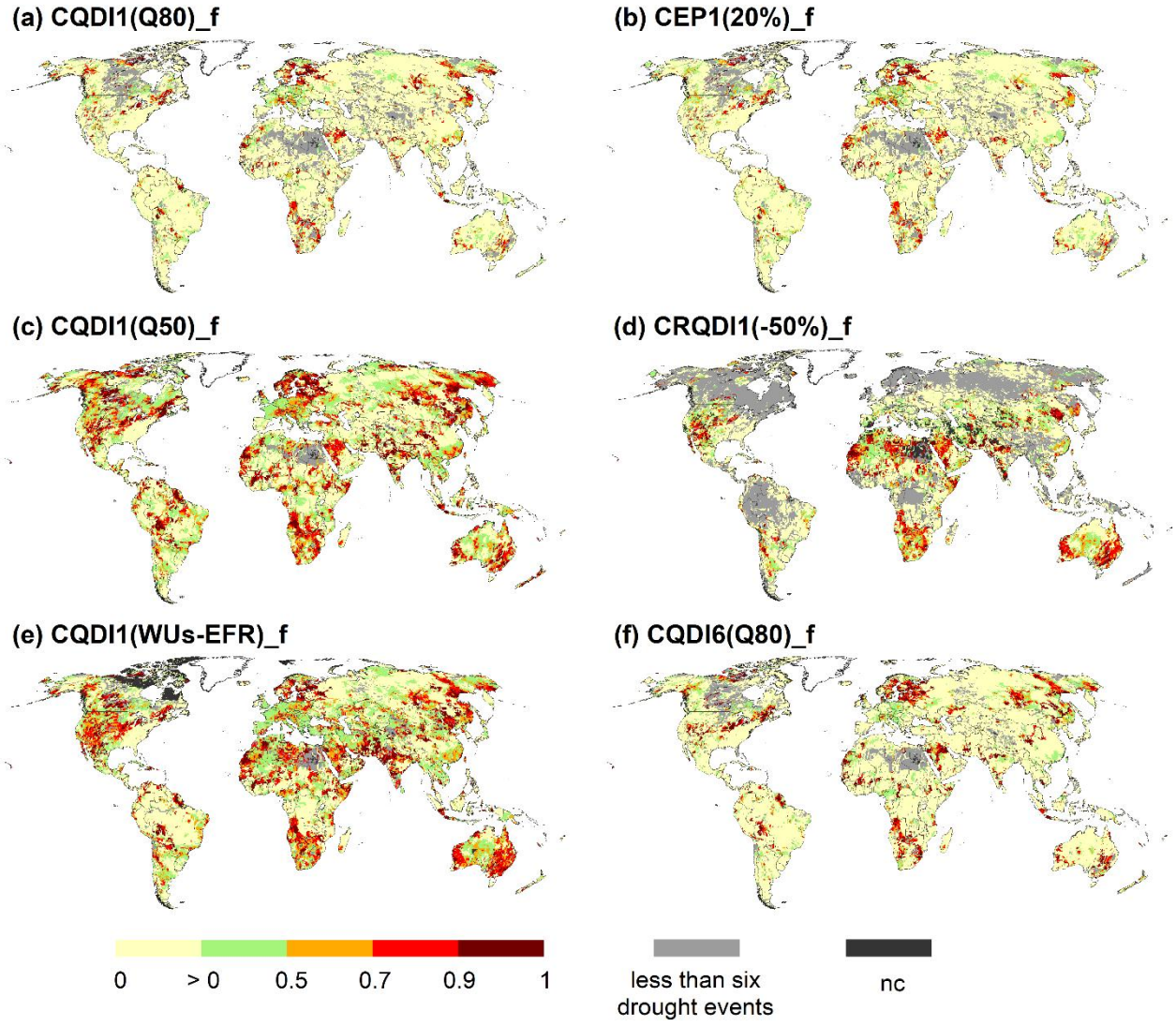


Figure 5: Probability of non-exceedance of drought events (level 2 in Fig. 1) in July 2003 for the cumulative indicators CQDI1(Q80)_f (a), CEP1(20%)_f (b), CQDI1(Q50)_f (c), CRQDI1(-50%)_f (d), CQDI1(WUs-EFR)_f (e), and CQDI6(Q80)_f (f) for

the reference period 1986-2015. A value of 0.8, for example, indicates that the cumulative anomaly or deficit, i.e., the severity up to this month, is higher than the severity of 80% of all drought events in the reference period. The probability of non-exceedance was not computed for grid cells shown in light grey, where less than six drought events were computed in the reference period (Sect. 2.43). "nc": not computable.

~~The relation between selected SDHIs from Figs. 4, 5 and 7 is compared for two illustrative grid cells that share the same CQDI Q80 value of 0.04 in March 2002 (Table 2). The grid cells are located in northern Italy (43.75° N, 11.25° E) and eastern Paraguay (24.25° N, 56.25° E) and are both characterized by low seasonal and high interannual streamflow variability.~~

~~Although SSI1 and EP1 are based on the same conceptual drought concept, they indicate different drought anomalies. The streamflow volume in March 2002 is the fourth smallest value in the Italian cell (rank 4, EP1=4/30=0.13) and the second smallest value in the Paraguayan cell (rank 2, EP1=2/30=0.07). However, since the slope at the lower bound of the ranked streamflow values in the Italian cell (not shown) is very small compared to the latter cell, the non-exceedance probability p of the fitted gamma distribution increases equally slowly, and the resulting p (and z score) is smaller than in the Paraguayan cell. In fact, the smallest SSI1 in March in the Italian grid cell is -1.76 and in the Paraguayan cell -1.56. Hence, a severe drought, usually defined below SSI1= 1.65, is never identified in March in the latter grid cells. Considering that the interannual variability is slightly higher in the Paraguayan cell, the results are in line with the hypothesis that standardized anomaly indicators may underestimate drought magnitude in such areas. RQDI1 on the other hand is lower in the latter cell reflecting a stronger streamflow deficit in March 2002. Moreover, drought magnitude in March 2002 for water users depending on reservoir storage (SSI12) is higher in the Paraguayan cell, indicating that the previous 12 months were relatively drier in the Paraguayan cell than in the Italian cell.~~

~~CEP1(20%)_f reveals that the rather strong streamflow anomaly in the Italian according to EP1 is only a peak within a moderate drought event as compared to the whole reference period. The value of 0.31 indicates that the drought severity up to March 2002 was exceeded by 69% of all drought events between 1986 and 2015. The low EP1 value of the Paraguayan cell is part of a more severe drought event that was exceeded by only 37% of all (completed) drought events. The higher drought magnitude in Paraguay according to RQDI1 corresponds, by chance, to a higher probability of non-exceedance of this drought event (CRQDI1(-50%)_f). This comparison underlines that indicators of drought magnitude are only suitable for assessing the current status of a drought event, but that they do not allow inferences about the status of the whole drought event compared to all other drought events of the reference period.~~

~~All severity indicators except CQDI1-Q80~~CQDI1(Q80) ~~indicate a stronger drought severity for the Paraguayan cell than for the Italian cell. Selection of Q50 instead of Q80 as threshold increases the cumulative water deficits from 4% to 36% (Paraguayan cell) and 25% (Italian cell) of mean annual streamflow volume. Selection of the sum of human surface water demand and environmental water demand as threshold indicates, with 56% and 51% of mean annual streamflow volume, respectively, even higher deficits since the onset of the drought event. Considering five indicators that express drought severity in terms of frequency, the non-exceedance frequency of drought severity in March 2002 ranges between 0.3 and 0.7 in the~~

Italian cell and between 0.6 and 0.8 in the Paraguayan cell (Table 2). In both grid cells, the indicators that do not assume habituation to interannual variability (CQDI1 Q50CQDI1(Q50)_f, CQDI1 WUsCQDI1(WUs)_EFR_f and CRQDI1(-50%)_f) show the highest severity.

Table 2: Comparison of SDHIs in March 2002 included in Figs. 4, 5 and 7 for a grid cell in northern Italy (43.75° N, 11.25° E) and eastern Paraguay (-24.25° N, -56.25° E).

Grid cell	Magnitude[1] (Current streamflow anomaly)				Severity[2] (Cumulative streamflow deficit)			Severity[3] (Probability (frequency) of non-exceedance)				
	SSI1	RQDI1	EP1	SSI12	CQDI1 Q80CQDI1(Q80)	CQDI1 Q50	CQDI1 WUs EFR	CQDI1 Q80_f	CQDI1 Q50_f	CQDI1 WUs EFR_f	CRQDI1(-50%)_f	CEP1 (20%)_f
Italy	-1.47[4]	-0.71	0.13[4]	-0.83	0.04	0.25	0.51	0.46	0.73	0.69	0.44	0.31
Paraguay	-1.08[5]	-0.78	0.07[5]	-1.32	0.04	0.36	0.56	0.57	0.83	0.80	0.83	0.63

[1]: Magnitude expressed as streamflow anomaly in March 2002 in units of standard deviation (SSI1) or empirical percentiles (EP1) with respect to mean March streamflow values during 1986-2015, relative deviation from mean March streamflow values (RQDI1), and streamflow anomaly averaged over April 2001 to March 2002 in units of standard deviation (SSI12) with respect to all April to March periods during 1986-2015.

[2]: Severity expressed as water volume deficit with respect to a threshold as a fraction of mean annual streamflow since drought onset until March 2002.

[3]: Severity expressed as probability (frequency) of non-exceedance in March 2002.

[4]: SSI1 = -1.47 is equivalent to Q93 and a return period of 14 years; EP1=0.13 is equivalent to a return period of 8 years.

[5]: SSI1 = -1.08 is equivalent to Q86 and a return period of 7 years; EP1=0.07 is equivalent to a return period of 15 years.

Although CEP1(20%) and CQDI1 Q80CQDI1(Q80) capture exactly the same drought signal for the grid cell in Northern Italy during the reference period (Fig. 8), the relative levels among the drought events differ. The two drought events in 1989 and 1990 have the highest severity levels according to both indicators. However, CQDI1 identifies the 1989 drought as the maximum event, while CEP1 detects the 1990 drought as the maximum event. This can be explained by the fact that the mean calendar month streamflow is lowest between June and October. Consequently, higher absolute streamflow deficit volumes (CQDI1) can build up during the 1989 drought ending in June 1989 than in the following drought event spanning over June to October 1990. Relative streamflow deficits on the other hand are larger for the latter. This is in line with the higher CQDI1 Q80CQDI1(Q80)_f value in March 2002 (0.46) compared to 0.31 for CEP1(20%)_f (Table 2), since the March 2002 event occurs outside the low-flow period. The short drought from June to August 2012 illustrates the difference between both

~~indicators even better. Among the twelve drought events, this drought event has only the second lowest drought severity according to CQDI1-Q80~~CQDI1(Q80)~~ f, but the 5th highest severity based on CEP1(20%) f. Consequently, when monitoring drought hazard, the severity of drought events during low-flow periods with high negative impacts is underestimated relative to other drought events based on frequency values of volume-based severity indicators. Of course, this is only true if monthly deficits are either not normalized (e.g. the low-flow index LFI, Cammalleri et al., 2016a) or normalized against mean annual streamflow volume (e.g. van Loon et al. (2014) and all CQDI1 variants in this paper) instead of mean monthly values.~~

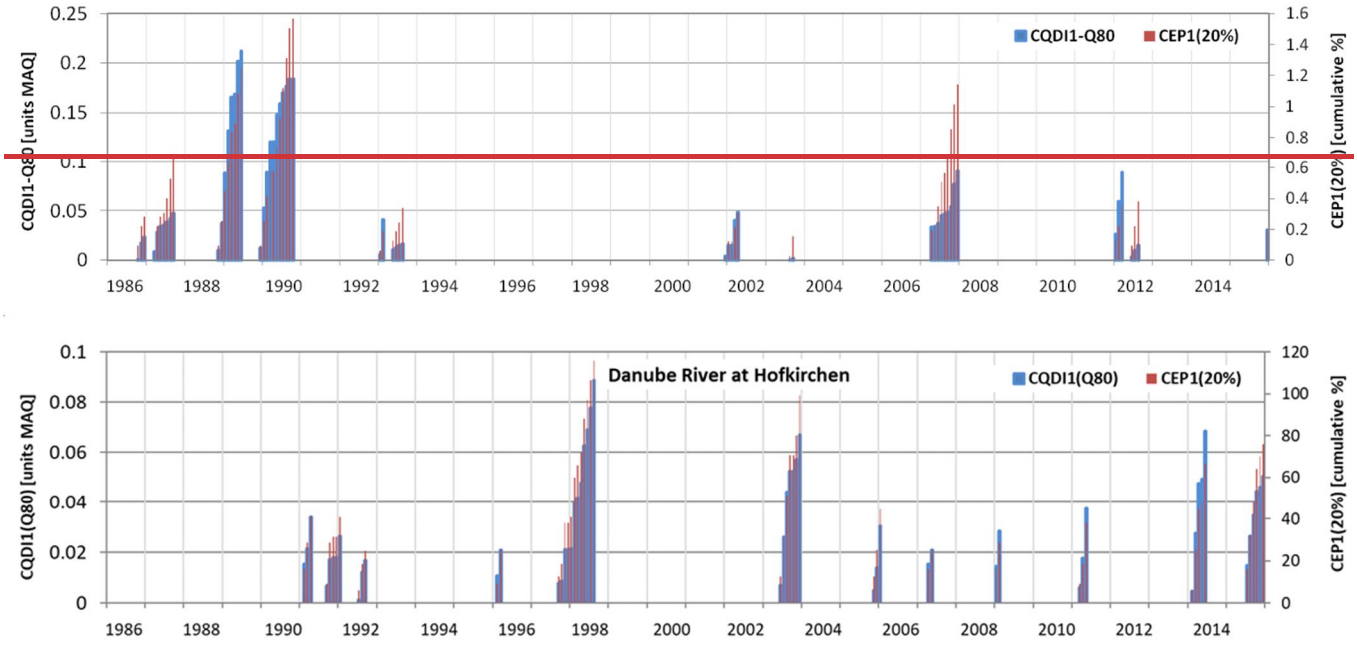
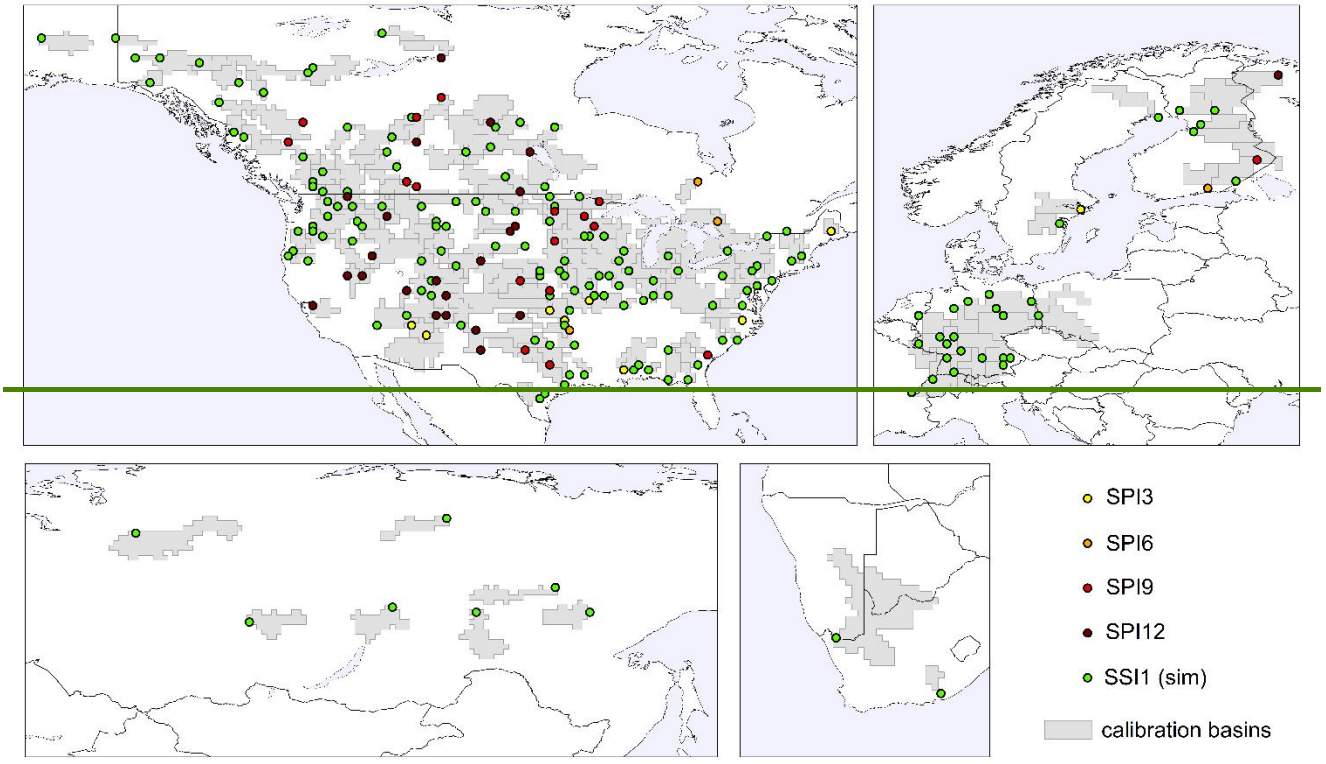


Figure ~~8~~6: Drought severity per month ~~(level 2)~~ during the reference period 1986-2015 at the Danube River, Hofkirchen, Germany,~~for a grid cell in Northern Italy (43.75° N, 11.25° E)~~ as indicated by ~~CQDI1-Q80~~CQDI1(Q80) (blue) and CEP1(20%) (red). MAQ: mean annual streamflow.

~~4.2.5 Suitability of SPIn to quantify streamflow drought hazard~~

~~To analyze if either simulated SSI1 (SSI1 (sim)) or SPIn is a better estimator of observed streamflow drought hazard, monthly time series of observed SSI1 (SSI (obs)) were correlated with five indicators applying the Pearson correlation: SSI1 (sim), SPI3, SPI6, SPI9, and SPI12. The analysis was limited to the 218 WaterGAP calibration stations with continuous time series of observed monthly streamflow during the reference period 1986-2015 for which all SPI variants were computable. Although longer averaging periods for SPI from 12 to 24 months are recommended for hydrological drought assessments (WMO and GWP, 2016), different studies at the global (Gevaert et al., 2018; Vicente-Serrano et al., 2012) and the regional scale (Yu et al., 2020; Huang et al., 2017; Barker et al., 2016) have demonstrated that shorter averaging periods often perform better in~~

~~estimating streamflow drought hazard. Using an ensemble of seven global land surface and hydrological models, Gevaert et al. (2018) found that the optimal SPI averaging period strongly varied among the models and with the season and climate regime. Different SPI variants from SPI1 to SPI24 were identified to correlate best with modeled SSI1 time series. In regional studies covering arid to humid climate and basin areas ≤ 10,000 km², SPI1 to SPI4 had the highest agreement with observed SSI1 values. Only in some humid basins underlain by productive aquifers, longer averaging periods from 6 to 19 months showed the best results. Basin size was found to be positively correlated with higher averaging periods (Yu et al., 2020). Moreover, shorter averaging periods performed better during spring and summer and longer averaging periods in autumn and winter (Huang et al., 2017). The strong influence of basin properties such as area and storage capacity as well as the climate regime was also discussed in van Loon (2015).~~



~~Figure 9: Ranking of the Pearson correlation coefficient r at 218 WaterGAP calibration stations between SSI1 (obs: observed streamflow) and each of the five indicators SSI1 (sim: simulated streamflow), SPI3, SPI6, SPI9 and SPI12 for the reference period 1986-2015.~~

~~Figure 9 depicts the 218 WaterGAP calibration stations and their basins located in North America, Northern and Western Europe, Russia, and South Africa. At each station, the drought hazard indicator that achieved the highest Pearson correlation coefficient is indicated. Overall, SSI1 (sim) performed best at 165 stations with a median correlation coefficient of 0.7,~~

followed by SPI12 at 23 stations. Among the 165 stations, the next highest correlation was achieved by SPI3 and SPI12 at 50 and 47 stations with median correlation coefficients of 0.65 and 0.46. The performance of WaterGAP is often lower in semi-arid basins, in particular where streamflow is highly altered by irrigation and man-made reservoirs, and in regions dominated by lakes. The SPI indicators outperform SSI1 (sim) only in smaller basins. Nonetheless, the total number of smaller basins (below 80,000 km²) where SSI1 (sim) is a better estimator of observed drought hazard still exceeds the number of basins where SPI variants perform better. Comparing the SPI variants, it would be expected that longer averaging periods should better capture the drought hazard in larger basins where drought propagation through the hydrological compartments is more delayed. Our analysis indicates a slight tendency that the 9- and 12-month averaging periods show better results in the larger basins, but this cannot be clearly deducted due to the limited sample size. In conclusion, drought hazard indicators based on modeled streamflow can often better quantify observed streamflow drought hazard than meteorological indicators, but this strongly depends on the goodness of fit after streamflow calibration.

5 Recommendations

5.1 General recommendations regarding the selection of drought hazard indicators

Drought hazard indicators serve to assess drought risk and to support drought management. Therefore, when deciding on which drought hazard indicator to use, it is first necessary to clearly define "the risk of what for whom" that is to be addressed by the indicator. Drought hazard indicators are risk system specific, and there is not one that fits all. As drought is conceptualized as both "less water than normal" and "less water than needed", the choice of indicator as well as the interpretation of the quantitative indicator values implicitly includes assumptions about what is normal and what is needed, i.e., to what amount of available water people or ecosystems are habituated to without suffering from drought. For example, is the risk bearer well adapted to the seasonal variability as well as the water availability that is exceeded in 8 out of 10 months per calendar month, even though the latter is only a small fraction of the mean or median water availability as is the case in dry regions of the globe? Then, an anomaly based indicator with the Q80 values of each calendar month as thresholds may be suitable. However, the spatial distribution of the values of such an anomaly indicator is only consistent and informative if it can be assumed that risk bearers everywhere on the map are habituated in the same way. Or is the risk bearer not well adapted to a high inter-annual variability and would suffer from even small reductions from mean/median streamflow? Then median values could be used as threshold. Alternatively, it could be assumed in such a case that the drought hazard increases with the percent reduction of water availability from average availability in a calendar month. Then, an indicator based on the relative deviation of the current water condition of the mean calendar month condition is informative. When selecting a drought hazard indicator, we recommend that the assumptions about the habituation of the risk bearer are made explicit first, based on knowledge about the risk system, which is then followed by the selection of a drought hazard indicator that fits to these assumptions.

~~When translating these conceptual drought definitions into operation drought hazard indicators, we recommend differentiating clearly drought magnitude indicators from drought severity indicators. Magnitude indicators with short averaging periods such as 1 month provide information on current, potentially extreme condition of the water flow or water storage under consideration. They should be used if the well-being of the risk bearer, e.g., river biota, strongly depends on the water condition at the specific time step of the analysis, e.g., the current month. As negative drought impacts are mostly assumed to increase with the length of the drought, however, severity indicators are often more informative than magnitude indicators as they quantify the cumulative magnitude since start of the drought. All drought magnitude indicators can be used to derive drought severity indicators. With exceptions, we recommend that drought severity at a certain point in time is expressed in terms of the probability/frequency of occurrence (return period) of a drought event with such a severity. These recommendations relate to any type of drought variable (precipitation, soil moisture, etc.) and spatial scale.~~

5~~.2~~ Recommendations for SDHIs in continental and global DEWS

Continental and global DEWS, which encompass near-real-time monitoring as well as seasonal forecasts, are to inform about drought hazards for diverse risk systems, which are characterized by different risk bearers (e.g., human water supply, river ecosystems), habituation, streamflow regimes, and water storage capacities. Therefore, a large-scale DEWS should provide data for a rather large number of drought hazard indicators together with a clear description of suitability for different risk systems including the underlying assumptions about habituation (or adaptation) of the risk bearer to the streamflow regime (Sect. 3.1). Then, end-users can select and combine several drought hazard indicators that are most informative. Table 3 lists the SDHIs that should be provided by large-scale DEWS, differentiating three risk groups and three main types of habituation~~for~~. In a DEWS, ~~1) human water supply from surface water and 2) river ecosystems, distinguishing intermittent and perennial streamflow regimes as well as low and large water storage capacities.~~drought magnitude indicators should be clearly differentiated from drought severity indicators (Sect. 3.2), and the specific suitability of each SDHI for different risk systems should be explained comprehensively.

~~(Meza et al., 2021; Meza et al., 2020)x~~

Table ~~3~~2: SDHIs for human water supply and river ecosystems ~~that should be provided by large-scale DEWS~~ for different risk groups. *Italic font*: Indicator assumes habituation to a certain degree of interannual variability (see Fig. A1b). **~~B~~bold font**: Indicator assumes the ability to fulfill seasonally varying demand for surface water abstractions and environmental flow. N~~n~~ormal font: Indicator assumes habituation to a certain reduction from mean monthly streamflow, and it is likely suitable for highly vulnerable systems with high interannual streamflow variability. All indicators assume habituation to the seasonality of streamflow.

~~, together with their suitability for drought risk assessment for 1) human water supply from surface water and 2) river ecosystems, distinguishing intermittent and perennial streamflow regimes as well as low and large water storage capacities~~

Font types indicate different types of habituation ~~Indicators of~~	~~Intermittent streamflow~~		~~Perennial streamflow~~	
	~~1) Water users without access to large reservoirs~~ ~~2) river ecosystems~~	~~1) Water users with access to large reservoirs~~	~~1) Water users without access to large reservoirs~~ ~~2) river ecosystems~~	~~1) Water users with access to large reservoirs~~
~~Magnitude~~	~~Return period based on EP1~~ ~~RQDI1~~	~~Return period based on EPn1~~ ~~RQDIn1~~	~~Return period based on EP1~~ ~~RQDI1~~	~~Return period based on EPn1~~ ~~RQDIn1~~
	~~In regions with (suspected) poor quality of hydrological model output, analyze SPEIn, in addition to streamflow indicators.~~			
~~Severity~~	~~CQDI1-Q80~~ ~~CQDI1-Q80_f~~	~~CQDIn-Q801~~ ~~CQDIn-Q80_f~~	~~CQDI1-Q802~~ ~~CQDI1-Q80_f2~~	~~CQDIn-Q801~~ ~~CQDIn-Q80_f~~
		CQDI1(Q80-HS) CQDI1(Q80-HS)~~_f~~		
		~~CEPn(20%)_f1~~ ~~CEP1(20%)-HS_f~~	~~CEP1(20%)_f2~~	
	~~CRQDI1(-50%)_f~~	~~CRQDIn(-50%)_f1~~ ~~CRQDI1(-50%)-HS_f~~	~~CRQDI1(-50%)_f~~	~~CRQDIn(-50%)_f1~~
	~~CQDI1-WUs-EFR~~ ~~CQDI1-WUs-EFR_f~~	~~CQDIn-WUs-EFR1~~ ~~CQDIn-WUs-EFR_f~~	~~CQDI1-WUs-EFR~~ ~~CQDI1-WUs-EFR_f~~	~~CQDIn-WUs-EFR1~~ ~~CQDIn-WUs-EFR_f~~

~~[1] n: For global-scale DEWS, an averaging period n of 6 or 12 months is suggested.~~

~~[2] CEP1(20%)_f preferred over CQDI-Q80 indicators.~~

~~*italics*: Indicator assumes habituation to a certain degree of interannual variability (see Fig. A1b).~~

bold: Indicator assumes the ability to fulfill seasonally varying demand for surface water abstractions and environmental flow.

normal font: Indicator assumes habituation to a certain reduction from mean monthly streamflow.

Risk group	Indicators of	
	Drought magnitude *	Drought severity
Water users without access to large reservoirs and r~~ well as R~~iver ecosystems	*Return period based on EP1*[1] RQDI1[2]	*CQDI1(Q80)*[3]*, CQDI1(Q80)_f*[4] *with streamflow deficit$_{m,y}$ = Q80$_m$ - Q$_{m,y}$* *CEP1(20%)_f*[5] *with percentile deficit$_{m,y}$ = P20 - EP1$_{m,y}$* CRQDI1(-50%)_f[6] with percent deficit$_{m,y}$ = -50% - RQDI1$_{m,y}$ **CQDI1(WUs-EFR)**[7]**, CQDI1(WUs-EFR)_f**[8] **with water deficit$_{m,y}$ = WUs$_m$ + EFR$_m$ - Q$_{m,y}$**~~*CQDI1(Q80)*[3]*, CQDI1(Q80)_f*[4]~~ ~~*CEP1(20%)_f*[5]~~ ~~CRQDI1(-50%)_f[6]~~ ~~**CQDI1(WUs-EFR)**[7]**, CQDI1(WUs-EFR)_f**[8]~~
Water users with access to ~~and~~ or downstream of large reservoirs	Same as in first ~~column~~row~~,~~ but with averaging periods of 6 and 12 months	Same as in first ~~column~~row~~,~~ but with averaging periods of 6 and 12 months~~diff~~
Reservoir managers **-***	Same as in first ~~column~~row~~,~~ but with averaging periods of 1, 6 and 12 month(s)	Same as in first ~~column~~row~~,~~ but with averaging periods of 1, 6 and 12 month(s) ~~suitableall suggested indicators for~~

* In regions with (suspected) poor quality of hydrological model output, analysis of SPEI6 and SPEI12 is suggested in addition to SDHIs.

** Reservoir managers should be informed to consider SDHIs of the grid cell~~s~~ that represent inflow into the reservoir.

[1] *EP1*: Empirical streamflow percentile per calendar month with an averaging period of one month, with 0<EP1≤100%. EP1 expresses the frequency of non-exceedance of the current streamflow. The return period, in years, is computed as 100/EP1. The lower the EP1 and the higher the return period, the higher the drought hazard.~~*EP1*: Empirical streamflow percentile per calendar month with an averaging period of one month, with 0<EP1≤1. EP1 expresses the frequency of exceedance of the current drought magnitude, while the inverse is the return period, in years. A low frequency value (or a high return period) indicates a high drought hazard~~

[2] RQDI1: Relative deviation of monthly streamflow from mean calendar month streamflow (MMQ), in percent. It is calculated as the difference between monthly streamflow and the respective MMQ, which is then divided by MMQ. The indicator is not computable in months with MMQ=0.

[3] *CQDI1(Q80)*: Cumulative volume-based streamflow deficit with an averaging period of one month divided by mean annual streamflow. A deficit occurs if monthly streamflow $Q_{m,y}$ falls below $Q80_m$ (the 20th percentile) of the long-term calendar month streamflow. Monthly deficits are accumulated for all drought months to obtain severity. A drought event starts with at least two consecutive months with a deficit, and it ends (deficit set to zero) if there are two consecutive months without a deficit (2mc: two- months criterion). Any streamflow surplus over Q80 in a single month between two deficit months does not decrease the cumulative deficit. To address flow intermittency, an existing drought continues during months where both Q80 and the current streamflow are zero. If Q80=0~~zero~~ and the current streamflow exceeds zero, the drought event ends. Hence, a drought can be prolonged, but never begin in calendar months with Q80=0. The indicator is expressed in units of mean annual streamflow.

[4] *CQDI1(Q80)_f*: The frequency of non-exceedance of drought events of a certain severity as quantified by CQDI1(Q80), with values between 0 and 1. A high frequency value indicates a high drought hazard. First, the partial duration series of drought events is derived based on the severities of all drought events of the reference period. Grid cells with less than six drought events are excluded. Second, the frequency of non-exceedance is quantified using the exponential cumulative distribution function. Preferably, the indicator should be expressed as return period T_r = ~~The return period is computed as~~ 1/(θ(1-~~CQDI1-Q80~~CQDI1(Q80)-_f)), with θ as the average number of drought events per year during the reference period.~~.~~

[5] *CEP1(20%)_f*: The frequency of non-exceedance of drought events of a certain severity (see above) as quantified by the cumulative percentile-based anomaly CEP1(20%). ~~The cumulative percentile based anomaly~~ The monthly percentile deficit is computed by subtracting the current streamflow percentile $EP1_{m,y}$ from the percentile threshold P20. Like CQDI1(Q80), CEP1(20%) allows an existing drought event to continue during months where both Q80 and the current streamflow are zero. The 2mc is also applied. ~~CEP1(20%) was computed in a similar way to CQDI1(Q80) using the 20th percentile per calendar month as threshold. It allows an existing drought event to continue during months where both Q80 and the current streamflow are zero. The two months criterion (see CQDI1(Q80)) is applied.~~ As the unit of CEP1(20%) (cumulative percent) is not informative, the indicator should be provided in frequency of non-exceedance or preferably as return period T_r = 1/(θ(1-CEP1(20%)_f)), with θ as the average number of drought events per year during the reference period.

[6] CRQDI1(-50%)_f: The frequency of non-exceedance of drought events of a certain severity (see above) as quantified by the~~The~~ cumulative relative streamflow deviation CRQDI1(-50%). The monthly percentile deficit is computed by subtracting the relative deviation of the current month $RQDI1_{m,y}$ from the threshold -50%.~~of monthly streamflow from mean calendar month streamflow (MMQ) with a threshold of RQDI1= 50%~~Months with MMQ=0, where the relative deviation is not computable, are defined to end a drought event assuming that people are habituated to zero streamflow in this month. The 2mc is also applied. ~~was derived like CQDI Q80. Months with MMQ=0, where RQDI1 is not computable, are defined to end a drought event.~~ As the unit of CRQDI1(-50%) (cumulative percent) is not informative, the indicator should be provided in frequency of non-exceedance or preferably return period T_r = 1/(θ(1-CRQDI1(-50%)_f)), with θ as the average number of drought events per year during the reference period.

[7] **CQDI1(WUs-EFR)**: Cumulative, volume-based water deficit with an averaging period of one month divided by mean annual streamflow. It is computed like CQDI1(Q80~~), but~~) but using the threshold mean monthly potential surface water abstraction WUs_m plus environmental flow requirement (EFR_m) per calendar month. EFR is assumed to be 80% of mean monthly naturalized streamflow Qnat per calendar month~~ such that 12 EFR values are obtained per grid cell~~. As EFR might never be fulfilled in case of strongly altered streamflow regimes, Qnat can be replaced by Qant implying that the river ecosystem has adapted to the altered streamflow conditions. WUs is the water demand from surface water bodies. The indicator is not computed in grid cells where mean annual WUs in the reference period is zero (approx. 9% of all grid cells excluding Greenland). The indicator is expressed in units of mean annual streamflow.

[8] **CQDI1(WUs-EFR)_f**: The frequency of non-exceedance of drought events of a certain severity (see above) as quantified by CQDI1(WUs-EFR)~~) (see description of CQDI1(Q80)_f~~). The indicator should be expressed as return period T_r = 1/(θ(1-CQDI1(WUs-EFR)_f)), with θ as the average number of drought events per year during the reference period.

To assess drought *magnitude*, we recommend using empirical percentiles and relative deviations to cover risk systems that are either habituated to a certain degree of interannual variability or to a certain reduction to mean calendar month streamflow. An averaging period of 1 month is suitable for river ecosystems and water users without access to large reservoirs, who depend on the currently available streamflow~~water as it flows down the river~~. Longer averaging periods of 6 or 12 months are suitable for~~in regions where~~ people ~~have~~ with access to or downstream of reservoir~~s storage~~ that are~~is~~ replenished during high-flow periods and that can alleviate short periods of below-normal streamflow. For reservoir managers, EP and RQDI with short and longer averaging periods (1, 6, and 12 months) are recommended for monitoring current reservoir inflow anomalies as well as reservoir storage anomalies (with different averaging periods depending on the storage capacity of the reservoir). Due to model uncertainties, time series of reservoir storage as simulated by WaterGAP should not be used for drought assessment. Importantly, reservoir managers should only consider SDHIs of the grid cell that represent inflow into the reservoir. This also applies if drought hazard for large lakes is analyzed by SHDIs.

-We favor empirical percentiles EP over SSI as the former are more transparent to end-users of a DEWS and do not entail uncertainties due to the fitting of a probability distribution. Moreover, application of one selected probability distribution function at large scales will always exclude many grid cells where the fitting is not possible. Here, other methods such as empirical percentiles would be required in any case. Expressing percentiles as return period (in years) may further increase the transparency of EP as end-users are accustomed to quantifying flood hazards by return periods. If the current streamflow is lower than the 30 values of the reference period, EP would only indicate a return period > 30 years (or a z score below -1.83), while SSI would indicate an extrapolated value, albeit with high uncertainty (Tijdeman et al., 2020). -Hence, 40 reference years should be used, if possible, to differentiate severe and extreme droughts with return periods of up to 40 years (equivalent to z = -1.96). In addition to the streamflow-based indicators, the standardized precipitation-potential evapotranspiration index SPEI (Vicente-Serrano et al., 2010) is suggested in regions with (suspected) poor quality of hydrological model output. Longer averaging periods of 6 and 12 months are recommended to consider the delayed response of streamflow to below-normal precipitation and potential evapotranspiration. However, it should be noted that meteorological indicators have limitations in describing hydrological drought processes (Haslinger et al., 2014; Blauhut et al., 2016; Laaha et al., 2017).

~~For all four risk systems in Table 3 (intermittent or perennial streamflow, both with and without access to large reservoirs), d~~Drought *severity* should be assessed with indicators that imply habituation to a certain degree of interannual variability (CEP(20%) ~~variants~~ and~~/or~~ CQDI(-Q80)) ~~variants)~~, to a certain reduction from mean monthly streamflow (CRQDI(-50%) ~~variants~~), and to the ability to fulfill seasonally varying human -water demand from surface water ~~abstractions~~ and environmental flow (CQDI(-WUs-EFR) ~~variants~~). Recommended averaging periods are the same as for magnitude indicators. With exceptions, we recommend that drought severity at a certain point in time is expressed in terms of the probability/frequency of non-exceedance (return period) of a drought event with such severity. These recommendations also relate to ~~to any type of drought~~ variable types other than streamflow (precipitation, soil moisture, etc.) and other spatial scales. In addition, the CQDI indicators should be provided in units of mean annual streamflow. ~~The severity indicators expressed in cumulative percent~~

~~(CEP and RQDI variants) should be provided as frequency of non-exceedance (denoted with suffix "f") as the informative value of a cumulative percentage is low.~~

CRQDI1(-50%) is preferred over CQDI1(Q50), which is based on a similar assumption about habituation since percent deviations are often applied in climate change impact studies and may thus be easier to grasp. Moreover, CQDI1(WUs-EFR) is preferred over CQDI1(WUs) since the environmental component of water demand should be considered in a DEWS. Regarding the percentile-based indicators CEP1(20%) and CQDI1(Q80%), the problem of flow intermittency is overcome by allowing an existing drought to continue during months where Q80 and the current streamflow are zero. CEP1 was found to be more sensitive to low-flow droughts than CQDI1, and it is therefore preferred over the latter if the risk system is more vulnerable to low-flow droughts than to high-flow droughts. ~~The severity indicators expressed in cumulative percent (CEP and RQDI variants) should be provided as frequency of non-exceedance (denoted with suffix "f") as the informative value of a cumulative percentage is low.~~

CQDI1(Q80-HS), conceptualized for risk systems with reservoirs, is not recommended due to the small impact of the HS criterion (Sect. 2.3.2) at the global scale.

~~Application of longer averaging periods of 6 or 12 months is recommended for all severity indicators in regions with large reservoirs where the impact of short term droughts below 6 months is probably low. In addition, the HS method (Sect. 2.2.3) is recommended in intermittent flow regimes with reservoirs for CQDI1, CRQDI1 and CEP1. The method allows existing high flow droughts to continue during low flow periods (defined as calendar months with Q80=0). The HS method follows the assumption that risk bearers in these regions cannot recover from high flow droughts during low flow periods such that drought severity should be kept at the initial level. CEP1(20%) HS_f (not assessed in this study) is computed like CQDI1(Q80 HS)~~CQDI1(Q80 HS)~~ f (Sect. 2.2.3) with potential drought prolongation in calendar months with Q80=0. CRQDI1(-50%) HS_f (not assessed in this study) allows an existing high flow drought to continue in calendar months where mean monthly streamflow MMQ is zero (and thus all 30 streamflow values in this calendar month). The CQDI variants are suitable for all risk systems as they inform end users of a DEWS about drought severity in units of absolute streamflow volume. CEP1 was found to be more sensitive to low flow droughts than CQDI1 (Fig. 8 and Sect. 4.2.4). In intermittent stream flow regimes without reservoirs, however, droughts are mostly detected during the high flow periods, and CEP1(20%)_f would not add value to a drought hazard assessment. In intermittent streams with reservoirs on the other hand, application of CEPn(20%)_f or CEP1(20%) HS_f is valuable, since these indicators quantify relative streamflow deficits, and the ranking of drought events by their severity is different from the ranking according to CQDI variants (Fig. 8). Regarding the risk for ecosystems or water supply in perennial rivers without large reservoirs, CEP1 is preferred over CQDI1 Q80~~CQDI1(Q80) due to the sensitivity of the former to low flow droughts. In perennial rivers with large reservoirs upstream, CQDIn Q80 is preferable.~~

According to Stahl et al. (2020), practitioners often use particular streamflow values rather than anomalies as the trigger for management actions. These practitioners could use forecasted RQDI1 as provided by the global-scale DEWS to determine whether this trigger will be reached by computing streamflow from RQDI1 and observed mean monthly streamflow.

6 Conclusions

This paper presents a new systematic approach for selecting global-scale streamflow drought hazard indicators (SDHIs) for monitoring drought risk for human water supply and river ecosystems in large-scale drought early warning systems (DEWS). The methodology replaces the conventional and imprecise classification into threshold-based and standardized indicators by a new classification scheme that distinguishes indicators pertaining to four indicator types by a) their inherent assumptions about the habituation of people and the ecosystem to the streamflow regime and b) their level of drought characterization, namely drought magnitude and drought severity~~, with the latter either in units of cumulative drought magnitude or in terms of frequency of occurrence.~~ The new scheme facilitates a better understanding of the information value of drought hazard indicators. It can support the development of a (large-scale) DEWS as well as water managers who rely on drought hazard indicators for their decision-making.

~~We applied the new classification scheme to a set of nine existing and three newly developed SDHIs for the reference period 1986-2015 using the global water resources and water use model WaterGAP 2.2d. Indicators of drought magnitude included SPI12, SPEI12, SSI1, SSI12, empirical percentiles EP1, and relative streamflow deviations from mean conditions, RQDI1 and RQDI12. Indicators of drought severity comprised the cumulative volume-based drought severity indicators CQDI1-Q50 and CQDI1-Q80 (with Q50 and Q80 as threshold), cumulative empirical percentiles CEP1(20%) (20th percentile as threshold), and cumulative relative deviations from mean conditions CRQDI1(-50%) (-50% as threshold). We developed a severity indicator for highly seasonal (HS) streamflow regimes with access to large reservoirs, CQDI1(Q80-HS) and two water deficit indicators, CQDI1-WUs and CQDI1-WUs-EFR, both considering mean monthly surface water use, and in case of the latter also mean monthly environmental flow requirements. These indicators cover several types of habituation to the streamflow regime comprising the habituation to a certain degree of interannual variability of streamflow, seasonality, a certain reduction from mean calendar month or mean annual streamflow, and being able to fulfill the demand for surface water abstractions and environmental flow. The comparison of indicators shows, for the first time explicitly for drought magnitude and severity, how conceptualization and selection of indicators can lead to very different spatial and temporal patterns of drought hazard. Using two example grid cells, the set of indicators resulted in a high range of non-exceedance frequencies (0.3-0.7 and 0.6-0.8) of drought severity of the same drought event. Indicators of drought magnitude are only suitable for assessing the current status of a drought event, but they do not allow inferences about the status of the whole drought event compared to all other drought events of the reference period.~~

~~A limited validation exercise revealed a tendency of WaterGAP 2.2d to overestimate observed Q80 between October and April, while Q80 during the low flow period in the Northern Hemisphere (May to September) is better captured. In a recent study comparing simulated and observed monthly streamflow and SSI3 at 183 stations worldwide, WaterGAP 2.2d model output showed a moderate agreement with observations. However, high agreement was identified at 25 stations in~~

~~Central and Eastern Europe, the United States, and South Africa. Moreover, the current study showed that SSI1 based on modeled streamflow often outperformed SPI with different averaging periods. However, this strongly depends on the goodness of fit after streamflow calibration. In uncalibrated basins, meteorological drought indicators should be used complementarily as proxies for hydrological drought hazard due to the uncertainty of modeled streamflow.~~

When providing drought hazard information in a global- or continental-scale DEWS, it is unknown which streamflow characteristics people and river ecosystems are locally accustomed to, and it is uncertain to what degree people have access to water stored in reservoirs. The suitability of hazard indicators is region- and risk system-specific (Blauhut et al., 2021) and can only be evaluated with local knowledge about the vulnerability of the system at risk. Therefore, a large-scale DEWS should provide data for a rather large number of drought hazard indicators that characterize the condition of various water flows (streamflow, actual evapotranspiration as a fraction of potential evapotranspiration) and water storage compartments (snow, soil, groundwater, lakes). Clear explanations for the end-users about the suitability of drought hazard indicators for specific risk systems need to be provided in DEWS. When selecting hazard indicators, we recommend that end-user make their assumptions about the habituation of the risk bearer explicit before selecting a drought hazard indicator that fits these assumptions. We suggest that f~~1) human water supply from surface water and 2) river ecosystems, distinguishing intermittent and perennial streamflow regimes as well as low and large water storage capacities.~~ uture studies analyze how well these hazard indicators, in combination with suitable vulnerability and exposure indicators, can estimate drought impacts in the targeted risk systems at regional or national scales.

Appendix

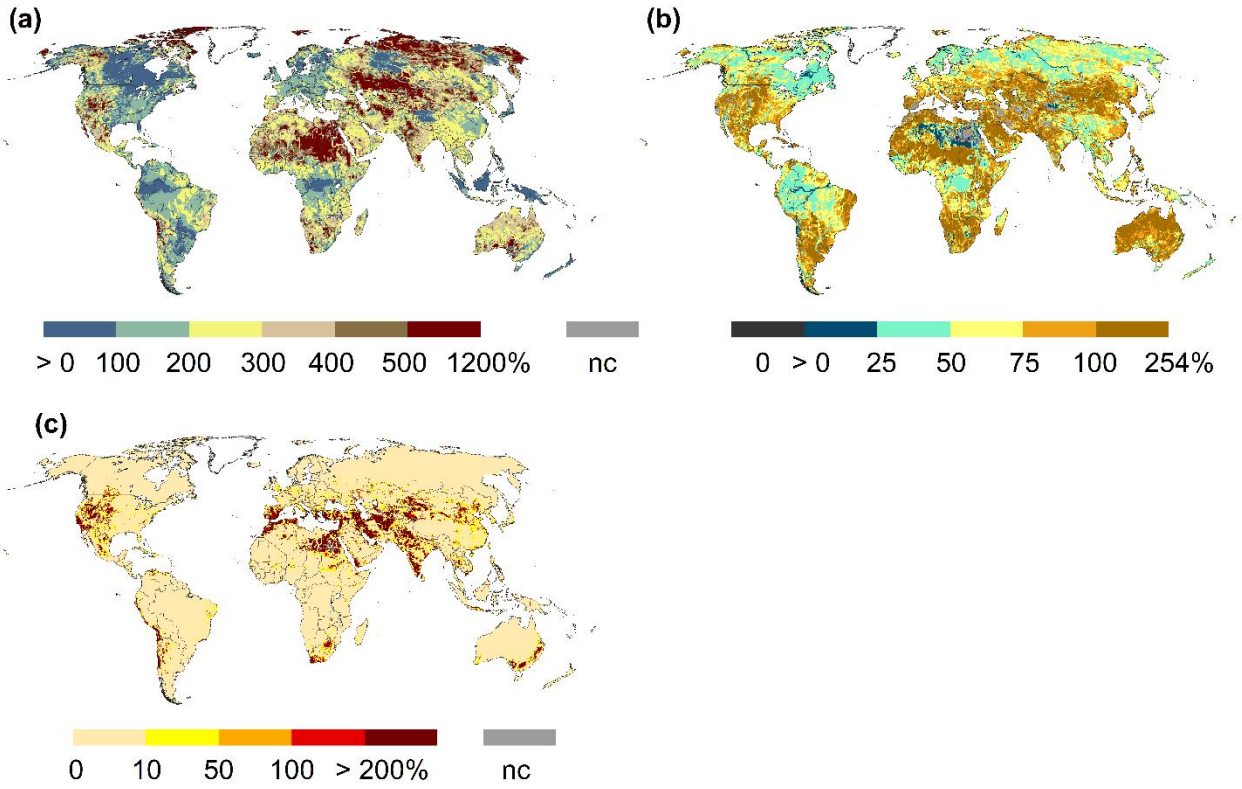


Figure A1: Seasonal streamflow variability indicated by the seasonal amplitude (Q in calendar month with highest mean monthly Q minus Q in calendar month with lowest mean monthly Q divided by MMQ (mean monthly Q over all calendar months)) (a), inter-annual streamflow variability indicated by the average of the 12 calendar month values of (Q20-Q80)/Qmean (b), and average of the 12 calendar month values of WUsmean/Qmean (c). All values in percent.

Data availability. WaterGAP 2.2d model output data used in this study are available at https://doi.org/10.1594/PANGAEA.918447 (Müller Schmied et al., 2021). The WaterGAP 2.2d source code is published at https://doi.org/10.5281/zenodo.6902111. The outputs from this study are available at https://zenodo.org/record/6647609 (Herbert and Döll, 2022). GRDC monthly streamflow data are available at: http://grdc.bafg.de (GRDC, 2019).

Supplement. The supplement related to this article is available online at: tbd

Author contributions. This manuscript was conceptualized by PD and CH. CH conducted the data analysis, visualization, and interpretation. The original draft was written by CH and revised by PD.

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

Acknowledgments. We thank the editor and ~~three~~ four anonymous referees for their thorough reviews and helpful suggestions for improving the manuscript. We thank Eklavyya Popat for computing the time series of SSI1 and SSI12.

Financial support. This research was part of the project GlobeDrought and has been supported by the German Federal Ministry of Education and Research (BMBF; grant no. 02WGR1457B) through its Global Resource Water (GRoW) funding initiative.

This open-access publication was funded by the Goethe University Frankfurt.

References

Abramowitz, M. and Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 55, 1965.

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., Knutson, C., Smith, K. H., Wall, N., Fuchs, B., Crossman, N. D., and Overton, I. C.: Drought indicators revisited: the need for a wider consideration of environment and society, WIREs Water, 3, 516–536, https://doi.org/10.1002/wat2.1154, 2016.

Barker, L. J., Hannaford, J., Parry, S., Smith, K. A., Tanguy, M., and Prudhomme, C.: Historic hydrological droughts 1891–2015: systematic characterisation for a diverse set of catchments across the UK, Hydrol. Earth Syst. Sci., 23, 4583–4602, https://doi.org/10.5194/hess-23-4583-2019, 2019.

Beguería, S.: Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, Journal of Hydrology, 303, 215–230, https://doi.org/10.1016/j.jhydrol.2004.07.015, 2005.

Blauhut, V., Stoelzle, M., Ahopelto, L., Brunner, M. I., Teutschbein, C., Wendt, D. E., Akstinas, V., Bakke, S. J., Barker, L. J., Bartošová, L., Briede, A., Cammalleri, C., Stefano, L. de, Fendeková, M., Finger, D. C., Huysmans, M., Ivanov, M., Jaagus, J., Jakubínský, J., Kalin, K. C., Krakovska, S., Laaha, G., Lakatos, M., Manevski, K., Neumann Andersen, M., Nikolova, N., Osuch, M., van Oel, P., Radeva, K., Romanowicz, R. J., Toth, E., Trnka, M., Urošev, M., Urquijo Reguera, J., Sauquet, E., Stevkova, S., Tallaksen, L. M., Trofimova, I., van Vliet, M. T. H., Vidal, J.-P., Wanders, N., Werner, M., Willems, P., and Živković, N.: Lessons from the 2018–2019 European droughts: A collective need for unifying drought risk management, 27 pp., 2021.

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., Stefano, L. de, and Vogt, J.: Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors, Hydrol. Earth Syst. Sci., 20, 2779–2800, https://doi.org/10.5194/hess-20-2779-2016, 2016.

Cammalleri, C., Barbosa, P., and Vogt, J. V.: Evaluating simulated daily discharge for operational hydrological drought monitoring in the Global Drought Observatory (GDO), Hydrological Sciences Journal, 65, 1316–1325, https://doi.org/10.1080/02626667.2020.1747623, 2020.

Cammalleri, C., Vogt, J., and Salamon, P.: Development of an operational low-flow index for hydrological drought monitoring over Europe, Hydrological Sciences Journal, 346–358, https://doi.org/10.1080/02626667.2016.1240869, 2016a.

Cammalleri, C., Micale, F., and Vogt, J.: A novel soil moisture-based drought severity index (DSI) combining water deficit magnitude and frequency, Hydrol. Process., 30, 289–301, https://doi.org/10.1002/hyp.10578, 2016b.

Corzo Perez, G. A., van Huijgevoort, M. H. J., Voß, F., and van Lanen, H. A. J.: On the spatio-temporal analysis of hydrological droughts from global hydrological models, Hydrol. Earth Syst. Sci., 15, 2963–2978, https://doi.org/10.5194/hess-15-2963-2011, 2011.

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrol. Earth Syst. Sci., 10, 535–552, https://doi.org/10.5194/hess-10-535-2006, 2006.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environmental Change, 23, 144–156, https://doi.org/10.1016/j.gloenvcha.2012.10.018, 2013.

GRDC: Global Runoff Data Centre, Federal Institute of Hydrology, Koblenz, Germany, 2019.

Griffiths, M. L. and Bradley, R. S.: Variations of Twentieth-Century Temperature and Precipitation Extreme Indicators in the Northeast United States, Journal of Climate, 20, 5401–5417, https://doi.org/10.1175/2007JCLI1594.1, 2007.

Hagenlocher, M., Meza, I., Anderson, C. C., Min, A., Renaud, F. G., Walz, Y., Siebert, S., and Sebesvari, Z.: Drought vulnerability and risk assessments: state of the art, persistent gaps, and research agenda, Environ. Res. Lett., 14, 83002, https://doi.org/10.1088/1748-9326/ab225d, 2019.

Haslinger, K., Koffler, D., Schöner, W., and Laaha, G.: Exploring the link between meteorological drought and streamflow: Effects of climate-catchment interaction, Water Resour. Res., 50, 2468–2487, https://doi.org/10.1002/2013WR015051, 2014.

Herbert, C. and Döll, P.: Streamflow drought hazard indicators for monitoring drought risk for human water supply and river ecosystems at the global scale, https://doi.org/10.5281/zenodo.6647609, 2022.

Heudorfer, B. and Stahl, K.: Comparison of different threshold level methods for drought propagation analysis in Germany, Hydrology Research, 48, 1311–1326, https://doi.org/10.2166/nh.2016.258, 2017.

Kumar, N. M., Murthy, C. S., Sesha Sai, M. V. R., and Roy, P. S.: On the use of Standardized Precipitation Index (SPI) for drought intensity assessment, Met. Apps, 16, 381–389, https://doi.org/10.1002/met.136, 2009.

Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., van Loon, A. F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J. H., and Wong, W. K.: The European 2015 drought from a hydrological perspective, Hydrol. Earth Syst. Sci., 21, 3001–3024, https://doi.org/10.5194/hess-21-3001-2017, 2017.

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.

Lloyd-Hughes, B.: The impracticality of a universal drought definition, Theor Appl Climatol, 117, 607–611, https://doi.org/10.1007/s00704-013-1025-7, 2014.

López-Moreno, J. I., Vicente-Serrano, S. M., Beguería, S., García-Ruiz, J. M., Portela, M. M., and Almeida, A. B.: Dam effects on droughts magnitude and duration in a transboundary basin: The Lower River Tagus, Spain and Portugal, Water Resour. Res., 45, https://doi.org/10.1029/2008WR007198, 2009.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, 1993.

Modarres, R.: Streamflow drought time series forecasting, Stoch Environ Res Ris Assess, 21, 223–233, https://doi.org/10.1007/s00477-006-0058-1, 2007.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geosci. Model Dev., 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

Nalbantis, I. and Tsakiris, G.: Assessment of Hydrological Drought Revisited, Water Resour Manage, 23, 881–897, https://doi.org/10.1007/s11269-008-9305-1, 2009.

Palmer, W. C.: Meteorological Drought, Research Paper No. 45, 45, 1965.

Popat, E. and Döll, P.: Soil moisture and streamflow deficit anomaly index: an approach to quantify drought hazards by combining deficit and anomaly, Nat. Hazards Earth Syst. Sci., 21, 1337–1354, https://doi.org/10.5194/nhess-21-1337-2021, 2021.

Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., Heim, R. R., Brewer, M. J., Svoboda, M., Westerhoff, R., van Dijk, A. I. J. M., Lloyd-Hughes, B., Pappenberger, F., Werner, M., Dutra, E., Wetterhall, F., Wagner, W., Schubert, S., Mo, K., Nicholson, M., Bettio, L., Nunez, L., van Beek, R., Bierkens, M., Goncalves, L. G. G. de, Mattos, J. G. Z. de, and Lawford, R.: Toward Global Drought Early Warning Capability: Expanding International Cooperation for the Development of a Framework for Monitoring and Forecasting, Bulletin of the American Meteorological Society, 94, 776–785, https://doi.org/10.1175/BAMS-D-11-00176.1, 2013.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.:

Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proceedings of the National Academy of Sciences of the United States of America, 111, 3262–3267, https://doi.org/10.1073/pnas.1222473110, 2014.

Richter, B. D., Davis, M. M., Apse, C., and Konrad, C.: A PRESUMPTIVE STANDARD FOR ENVIRONMENTAL FLOW PROTECTION, River Res. Applic., 28, 1312–1321, https://doi.org/10.1002/rra.1511, 2012.

Satoh, Y., Shiogama, H., Hanasaki, N., Pokhrel, Y., Boulange, J. E. S., Burek, P., Gosling, S. N., Grillakis, M., Koutroulis, A., Müller Schmied, H., Thiery, W., and Yokohata, T.: A quantitative evaluation of the issue of drought definition: a source of disagreement in future drought assessments, Environ. Res. Lett., 16, 104001, https://doi.org/10.1088/1748-9326/ac2348, 2021.

Sharma, T. C. and Panu, U. S.: Predicting return periods of hydrological droughts using the Pearson 3 distribution: a case from rivers in the Canadian prairies, Hydrological Sciences Journal, 60, 1783–1796, https://doi.org/10.1080/02626667.2014.934824, 2015.

Shukla, S. and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, Geophys. Res. Lett., 35, https://doi.org/10.1029/2007GL032487, 2008.

Smakhtin, V.: Low flow hydrology: a review, Journal of Hydrology, 240, 147–186, https://doi.org/10.1016/S0022-1694(00)00340-1, 2001.

Spinoni, J., Barbosa, P., Jager, A. de, McCormick, N., Naumann, G., Vogt, J. V., Magni, D., Masante, D., and Mazzeschi, M.: A new global database of meteorological drought events from 1951 to 2016, Journal of Hydrology: Regional Studies, 22, 100593, https://doi.org/10.1016/j.ejrh.2019.100593, 2019.

Spinoni, J., Muñoz, C. A., Masante, D., McCormick, N., Vogt, J. V., and Barbosa, P.: European Drought Observatory User Meeting: JRC Conference and Workshop Reports, 2018.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., van Loon, A. F., and Stahl, K.: Candidate Distributions for Climatological Drought Indices (SPI and SPEI), Int. J. Climatol., 35, 4027–4040, https://doi.org/10.1002/joc.4267, 2015.

Stahl, K., Vidal, J.-P., Hannaford, J., Tijdeman, E., Laaha, G., Gauster, T., and Tallaksen, L. M.: The challenges of hydrological drought definition, quantification and communication: an interdisciplinary perspective, Proc. IAHS, 383, 291–295, https://doi.org/10.5194/piahs-383-291-2020, 2020.

Steinemann, A., Iacobellis, S. F., and Cayan, D. R.: Developing and Evaluating Drought Indicators for Decision-Making, Journal of Hydrometeorology, 16, 1793–1803, https://doi.org/10.1175/JHM-D-14-0234.1, 2015.

Tijdeman, E., Stahl, K., and Tallaksen, L. M.: Drought Characteristics Derived Based on the Standardized Streamflow Index: A Large Sample Comparison for Parametric and Nonparametric Methods, Water Resour. Res., 56, https://doi.org/10.1029/2019WR026315, 2020.

UNECE: Policy Guidance Note on the Benefits of Transboundary Water Cooperation: Identification, Assessment and Communication,

unece.org/fileadmin/DAM/env/water/publications/WAT_Benefits_of_Transboundary_Cooperation/ECE_MP.WAT_47_PolicyGuidanceNote_BenefitsCooperation_1522750_E_pdf_web.pdf, 2015.

van Huijgevoort, M. H. J., Hazenberg, P., van Lanen, H. A. J., and Uijlenhoet, R.: A generic method for hydrological drought identification across different climate regions, Hydrol. Earth Syst. Sci., 16, 2437–2451, https://doi.org/10.5194/hess-16-2437-2012, 2012.

van Huijgevoort, M., van Lanen, H., Teuling, A. J., and Uijlenhoet, R.: Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge, Journal of Hydrology, 512, 421–434, https://doi.org/10.1016/j.jhydrol.2014.02.060, 2014.

van Lanen, H. A. J.: Drought propagation through the hydrological cycle, 2006.

van Lanen, H., Vogt, J. V., Andreu, J., Carrão, H., Stefano, L. de, Dutra, E., Feyen, L., Forzieri, G., Hayes, M., Iglesias, A., Lavaysse, C., Naumann, G., Pulwarty, R., Spinoni, J., Stahl, K., Stefanski, R., Stilianakis, N., Svoboda, M., and Tallaksen, L. M. (Eds.): Climatological risk: droughts: Poljanšek, K., Marín Ferrer, M., De Groeve, T., Clark, I. (Eds.). Science for disaster risk management 2017: knowing better and losing less., 556 pp., 2017.

van Loon, A. F., Tijdeman, E., Wanders, N., van Lanen, H. A. J., Teuling, A. J., and Uijlenhoet, R.: How climate seasonality modifies drought duration and deficit, J. Geophys. Res. Atmos., 119, 4640–4656, https://doi.org/10.1002/2013JD020383, 2014.

van Loon, A. F., van Huijgevoort, M. H. J., and van Lanen, H. A. J.: Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, Hydrol. Earth Syst. Sci., 16, 4057–4078, https://doi.org/10.5194/hess-16-4057-2012, 2012.

van Loon, A. F.: Hydrological drought explained, WIREs Water, 2, 359–392, https://doi.org/10.1002/wat2.1085, 2015.

van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., van Dijk, A. I. J. M., Tallaksen, L. M., Hannaford, J., Uijlenhoet, R., Teuling, A. J., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., and van Lanen, H. A. J.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, Hydrol. Earth Syst. Sci., 20, 3631–3650, https://doi.org/10.5194/hess-20-3631-2016, 2016.

van Oel, P. R., Martins, E. S. P. R., Costa, A. C., Wanders, N., and van Lanen, H. A. J.: Diagnosing drought using the down-streamness concept: the effect of reservoir networks on drought evolution, Hydrological Sciences Journal, 63, 979–990, https://doi.org/10.1080/02626667.2018.1470632, 2018.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, Journal of Climate, 23, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., and Baillon, M.: Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, 20 pp., 2009.

Vincent, L. A. and Mekis, É.: Changes in Daily and Extreme Temperature and Precipitation Indices for Canada over the Twentieth Century, Atmosphere-Ocean, 44, 177–193, https://doi.org/10.3137/ao.440205, 2006.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, https://doi.org/10.1002/2014WR015638, 2014.

Wilhite, D. and Glantz, M.: Understanding the drought phenomenon: the role of definitions, Water International, 10, 111–120, https://doi.org/10.1080/02508068508686328, 1985.

WMO and GWP: Handbook of Drought Indicators and Indices (M. Svoboda and B.A. Fuchs). Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2. Geneva., 2016.

Yevjevich, V.: An objective approach to definitions and investigations of continental hydrological droughts, Hydrology Papers Colorado State University, 1967.

Yihdego, Y., Vaheddoost, B., and Al-Weshah, R. A.: Drought indices and indicators revisited, Arab J Geosci, 12, https://doi.org/10.1007/s12517-019-4237-z, 2019.

Zaidman, M. D., Rees, H. G., and Young, A. R.: Spatio-temporal development of streamflow droughts in north-west Europe, Hydrol. Earth Syst. Sci., 6, 733–751, https://doi.org/10.5194/hess-6-733-2002, 2002.