Received: 12 December 2017 | Revised: 13 March 2018 | Accepted: 14 March 2018

DOI: 10.1002/wat2.1288

WILEY WIREs WATER

ADVANCED REVIEW

Robustness of hydroclimate metrics for climate change impact research

Marie Ekström[1,2] | Ethan D. Gutmann[3] | Robert L. Wilby[4] | Mari R. Tye[5] | Dewi G.C. Kirono[6]

[1]Land and Water, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, Australia

[2]School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK

[3]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado

[4]Department of Geography, Loughborough University, Loughborough, UK

[5]Mesoscale and Microscale Meteorology Laboratory, National Center for Atmospheric Research, Boulder, Colorado

[6]Climate Science Centre, Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Aspendale, Australia

Correspondence
Marie Ekström, School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK.
Email: ekstromm@cardiff.ac.uk

Funding information
NSF EaSM3, Grant/Award Number: 1419563; National Science Foundation; Cardiff University Water Research Institute; School of Earth and Ocean Sciences of Cardiff University; Australian Government's National Environmental Science Programme

Metrics based on streamflow and/or climate variables are used in water management for monitoring and evaluating available resources. To reflect future change in the hydrological regime, metrics are estimated using climate change information from Global Climate Models or from stochastic time series representing future climates. Whilst often simple to calculate, many metrics implicitly represent complex physical process. We evaluate the scientific validity of metrics used in a climate change context, demonstrating their use to reflect aspects of timing, magnitude, extreme values, variability, duration, state, system services, and performance. We raise awareness about the following generic issues (a) *formulation*: metrics often assume stationarity of the input data, which is invalid under climate change; and do not always consider potential changes to seasonality and the relevance of the temporal window used for analysis; (b) *climate change input data*: how well are the physical processes relevant to the metric represented in the climate change input data; what is the impact of bias correction on relevant spatial and temporal scale dependencies and relevant intervariable dependencies; how realistic are the data in representing sequencing of events and natural variability in large-scale ocean–atmosphere systems; and (c) *decision-making context*: are rules and values that frame the decision-making process likely to remain constant or change in a future world.

If critical climate or hydrological processes are not well represented by the metric constituents, these indices can be misleading about plausible future change. However, knowledge of how to construct a robust metric can safeguard against misleading interpretations about future change.

This article is categorized under:
Science of Water > Methods
Science of Water > Water and Environmental Change
Science of Water > Water Extremes

KEYWORDS
climate change, hydroclimate, metrics

1 | INTRODUCTION

Metrics are widely used to characterize properties of the hydrological regime, such as variations in river discharge to support water management (Kennard, Mackay, Pusey, Olden, & Marsh, 2010). From a planning perspective, metrics can simplify analysis of long-term trends in flow characteristics and improve understanding of the impacts of streamflow variability on water resources (Cherkauer & Sinha, 2010; Rouge & Cai, 2014). Metrics can clarify relationships between different variables

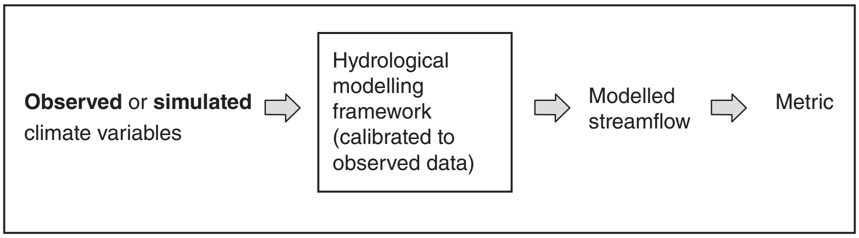

FIGURE 1 Most metrics surveyed here are derived from streamflow estimated by hydrological modeling. In most hydrological impact modeling studies, simulated present and future climate variables are used as input to a hydrological model, which is typically calibrated using historic hydrological records

or processes, such as the rainfall-runoff ratio. Hydroclimate metrics can also be used to support water planning decisions or climate adaptation (Fatichi, Rimkus, Burlando, Bordoy, & Molnar, 2015). For example, Lawson, Fryirs, Lenz, and Leishman (2015) used average magnitude of daily streamflow within each season to link diversity in riparian ecosystems to flooding and seasonal water availability in southeast Australia. By comparing magnitudes and variability in metrics estimated from historical records with those based on climate change information, situations leading to service failure can be identified. Such insights can aid formulation of alternative operational strategies, upgrading and retrofitting of existing water supply infrastructure, or inform the design standards needed for new, long-lived water infrastructure.

Here, we seek to raise awareness about which metrics, if any, can be most confidently deployed in climate change impact research given the inherent uncertainties in the modeling chain between Global Climate Model (GCM) to impact model (Chiew et al., 2010; Clark et al., 2016; Teng, Vaze, Chiew, Wang, & Perraud, 2012) and our often limited understanding of current hydrological systems due to insufficient observational data.

Our call for increased awareness stems from the apparent ease by which metrics can be calculated, without the subsequent user of the metrics necessarily appreciating the potential for incorrect metric formulation, and consequential scope for maladaptation in the future. This challenge of satisfying information needs whilst adhering to scientific rigor led Wilby (2010), to propose several principles for users of climate model output in hydrological applications. Good practice in this regard includes recognizing and handling uncertainties in observations (e.g., Newman et al., 2015) and having realistic expectations about what information can be obtained from climate models. The latter is central to our evaluation of hydrological metrics for adaptation planning, which often relies heavily on climate model information.

2 | OVERVIEW OF HYDROCLIMATE METRICS

This review focuses on metrics (numerical indices) derived from hydrological streamflow observations and model simulations that describe various aspects of the hydroclimate relevant to water resources management (Figure 1). We further consider metrics based on climate variables that serve as indicators of regional climate trends and can provide contextual information about long-term variability in the hydroclimate. For example, the "annual precipitation total" metric was used to analyze trends and change points in streamflow regimes and water quality of Lake Winnipeg, North America (Ehsanzadeh, van der Kamp, & Spence, 2012). Similarly, the 100-year 24-hour rainfall (99th quantile) was used to characterize rare events relevant to storm-water design in Wisconsin, United Kingdom (Schuster, Potter, & Liebl, 2012).

The metrics presented here are collected from a literature search of studies where metrics are applied in a climate change context, to estimate impacts or support adaptation planning activities. Thus, this is not an exhaustive list of all the metrics that could possibly be used in climate change impact research. For example, metrics of economic performance in water resource management (Hurd & Rouhi-Rad, 2013) or water quality (von der Ohe et al., 2007) could be considered relevant, but are not considered here as they involve additional discipline-specific modeling and uncertainties. Model optimization is also not covered, but we do acknowledge that the skill of hydrological models depends on the choice of metric (i.e., objective function) linked to the parameter set, model structure (Fowler, Peel, Western, Zhang, & Peterson, 2016), and information content of the calibration and validation data (Broderick, Matthews, Wilby, Bastola, & Murphy, 2016; Vaze et al., 2010). This is particularly important in a climate change context because models are often the basis for future assessments, and any metric that is not well simulated by a model in the current climate is unlikely to have a reliable change signal in that model. For instance, a model calibrated to Nash-Sutcliffe efficiency will likely underestimate the variability in streamflow (Gupta, Kling, Yilmaz, & Martinez, 2009), and may not provide reliable guidance on the changes in, for example, extreme flood related metrics.

To structure our critique, we employ a categorization that groups metrics which describe similar properties. These groups cover:

- *Timing* of the seasonal flow-regime or hydrograph, such as when streamflow minima or maxima typically occur.
- *Magnitude* of discharges under average, low- or high-flow conditions, or when integrated over annual, seasonal, or peak daily flows.

- *Extreme values* of the magnitude or frequency of rare hydrological events.
- *Variability* of flow over various timescales, spanning flashiness, or counts of low- and high-flows over annual to multidecadal periods.
- *Duration* of hydrological events of interest, including persistence of low- or high-flow conditions.
- *State* of natural or managed systems described by water-balance, drought or aridity terms, including the runoff ratio or standardized precipitation index.
- *Service and performance* of a particular service or system, such as the "level of service" metric or the "water supply capacity" index.

2.1 | Timing

This group of metrics is used to identify the onset or completion of a process (such as snowmelt), the influence of a seasonally dominant weather pattern, or when a certain magnitude of runoff has passed the outlet of a catchment (Table 1, section Timing). They can reveal changes in seasonal storage, or risks to water supply, indicating a need to identify strategies for how and when existing resources are used, or whether other sources may need to be deployed.

The temporal unit for several metrics is the "water year". However, the start date for the 12-month period may vary regionally, reflecting the regional hydroclimate regime or harmonization with management cycles of water regulation. For example, the water year in the United States and the United Kingdom runs from 1 October to 30 September the following year (Reed, 1999; USGS, 2016), ensuring capture of snow falling in late autumn to early spring. In Australia, the water year starts on 1 July in-line with allocation of water entitlements and water market reports (Water Act 2007).

As with all metrics, it is important to select timing metrics that are resilient to random variability in flows. For example, the runoff centroid timing (CT) is typically used in regions where snowmelt is a significant component of river flow (Equation (1) and Table 1). Court (1962) used the index to assess the impact of delayed snowmelt in the Sierra Nevada on water management. The author asserted that this measure (referred to as "half-flow") is preferred to instantaneous maxima, which reflects peak events, as it better represents the accumulated flow within the water year. In contrast, the timing of instantaneous peak flow can vary depending on the coincidence of a specific rain event or warm spell with the period of peak snowmelt. Null et al. (2010) used the CT to assess watershed resilience to climate change in the Sierra Nevada. Their intent was to determine the date at which half the annual runoff at the outlet of each catchment has passed:

$$\mathrm{CT} = \frac{\sum t_i q_i}{\sum q_i} \text{ for } i = 1\ldots52 \tag{1}$$

where t_i is the time in weeks from the beginning of the water year and q_i is the streamflow for week i. Others used the metric to: demonstrate sensitivity of catchment behavior to hydrologic modeling structure and parameter choice under projected climate change in Colorado, United States (Mendoza et al., 2016); detect and attribute trends in western United States streamflow (here defined as the date of the calendar year when 50% of the water year flow has passed (Hidalgo et al., 2009); evaluate how well the hydrological regime is modeled in smaller streams in the Pacific Northwest United States (Wenger et al., 2010).

While CT is more stable than some metrics, Dery et al. (2009) suggests that where late season precipitation and/or glacial melt may contribute to discharge, CT-type metrics can lead to misleading results. For example, greater accumulation of snow may appear as a shift towards a later spring freshet, although the timing of spring snowmelt may not have changed. The authors suggested normalizing successive 5-day means of streamflow to study shifts in timing of flow. However, Hidalgo et al. (2009) claim that simpler metrics can still be meaningful, noting that for the western United States there is little correlation between annual streamflow volume and the centre timing metric. To detect changes in the timing of snow melt (SM) onset, Clow (2010) defined onset as the beginning of a 5-day period in which measured snow in the basin declined by 2.5 cm, a threshold set high enough to avoid sublimation effects. The author then used the day on which 20% of the annual streamflow volume had accumulated to identify the onset of the melt pulse in streamflow. However, these thresholds will be highly basin dependent. For example, a 20% threshold may be too low in basins with greater mid-winter streamflow. That metrics can be more or less robust in different regions, demonstrates the importance of understanding the physical meaning of a metric in each specific application.

Other metrics capture the timing of a defined flow event such as the Julian day of annual minimum (JMinF) or annual maximum (JMaxF) used by Zhang et al. (2016) to classify the regional streamflow regime in arid and semi-arid regions of northwest China. They found that JMaxF was preferred to JMinF, perhaps due to poorer simulation of low-flow events due to groundwater interplay and glacial melt processes. However, by relying on minimum and maximum values, these metrics may exhibit a high degree of interannual variability. For continental Europe, Blöschl et al. (2017) focused on the dates of

TABLE 1 Metric summary

Metric	Brief description	Unit	Reference
Timing			
CT	Runoff centroid, centre timing	Julian week in water year/date	Null, Viers, and Mount (2010) Hidalgo et al. (2009)
SM onset	Snow melt onset	cm depth/5 days	(Clow, 2010)
JMinF	Julian day of annual minimum	Julian day	Zhang, Shao, Zhang, Zhai, and She (2016)
JMaxF	Julian day of annual maximum	Julian day	Zhang et al. (2016)
Magnitude			
Q_0, Q_{10}, Q_{20}, Q_{30}, Q_{40}, Q_{50}, Q_{60}, Q_{70}, Q_{80}, Q_{90}, Q_{100}	Annual deciles of daily flow	Not given, but indicators are estimated from daily streamflow	Rouge and Cai (2014)
Q_0, Q_{25}, Q_{50}, Q_{75}, Q_{100}	Seasonal quartiles of daily flow	As above	Rouge and Cai (2014)
MMDF	Middle flow (25th–75th percentile)	m^3/s	Zhang et al. (2016)
Low75, Q95	Low-flow discharge (set percentile threshold)	log m^3/s	Zhang et al. (2016), Wade, Rance, and Reynard (2013).
Hig25	High-flow discharge (set percentile threshold)	m^3/s	Zhang et al. (2016)
MinF or minimum flow	Mean annual minimum flow	log m^3/s	Zhang et al. (2016), Cherkauer and Sinha (2010)
MaxF	Mean annual maximum flow	m^3/s	Zhang et al. (2016) Cherkauer and Sinha (2010)
FMM	Flow duration curve median	log m^3/s	Mendoza et al. (2016)
Peak flow	Average annual or seasonal peak daily flow	m^3/s	Cherkauer and Sinha (2010)
Mean flow	Mean annual and seasonal flow	For example, mm/day	Cherkauer and Sinha (2010), Rouge and Cai (2014), Wenger, Luce, Hamlet, Isaak, and Neville (2010), Kirono et al. (2014)
High-flow sums	Average annual or seasonal cumulative sum of flows above a selected daily flow level	km^3	Cherkauer and Sinha (2010)
Low-flow deficit	Average annual or seasonal cumulative sum of flows below a selected daily flow level	km^3	Cherkauer and Sinha (2010)
Q_{Feb}	Proportion of flow in February	Dimensionless	Booker and Woods (2014)
Q_{MALF}, 7-day min	7-day minimum flow	For example, mm/day	Booker and Woods (2014), Rouge and Cai (2014)
Max flow	Maximum daily flow	m^3/s	Ehsanzadeh et al. (2012)
Q05	Fifth percentile level of normalized flow		Pechlivanidis and Arheimer (2015)
HSPeak	Mean magnitude of high spells above the 95th percentile on the flow duration curve (standardized by mean daily flow)	Dimensionless	Lawson et al. (2015)
FLV	Flow duration curve low-segment volume	log m^3/s	Mendoza et al. (2016)
T_{Qmean}	Fraction of time that daily streamflow exceeds mean streamflow for each year	%	Cherkauer and Sinha (2010)
MDFMDFSpring[a]	Average for mean daily flow for spring (standardized by overall mean daily flow)	Dimensionless	Lawson et al. (2015)
Extreme values			
$30Q_{20}$	30-day, 20-year low-flow event	m^3/s	Maldonado and Moglen (2013)
3D30Y	3-day runoff volume during a 30-year runoff simulation	Not defined	Brekke et al. (2009)
W99 and W95	Number of days in winter in the top First and fifth percentile annual flows	Days	Wenger et al. (2010)
S95	Number of days in summer in the top fifth percentile annual flows	Days	Wenger et al. (2010)
W1.5 and W2	Probability for a 1 or 2-year flow event to occur in winter	Dimensionless	Wenger et al. (2010)
S10, S20	Number of days in the summer in which flows were less than 10/20% of mean annual flow	Days	Wenger et al. (2010)
7Q10	Seven-day average low-flow with a 10-year return period metric	Not defined	Wenger et al. (2010)
POT1–5	Peak over threshold sampling	Average number of peaks per year	Rouge and Cai (2014)

TABLE 1 (Continued)

Metric	Brief description	Unit	Reference
Variability			
FMS	Flow duration curve mid-segment slope	log m^3/s	Mendoza et al. (2016)
m_{FDC}	Slope of flow duration curve	Dimensionless	Pechlivanidis and Arheimer (2015), Viglione et al. (2013)
CVDF	CV of daily flow	Dimensionless	Zhang et al. (2016)
LowC75, low-flow count	Low-flow spell count (< percentile threshold)	Dimensionless	Zhang et al. (2016), Cherkauer and Sinha (2010)
HigC25	High-flow spell count (> percentile threshold)	Dimensionless	Zhang et al. (2016), Cherkauer and Sinha (2010)
CVAnnHSPeak	CV of all years' mean high-spell magnitude	Dimensionless	Lawson et al. (2015)
CVAnnHSNum	CV of all years' number of high spells	Dimensionless	Lawson et al. (2015)
CVMDFSpring[b]	CV of mean daily flow for spring	Dimensionless	Lawson et al. (2015)
MDFAnnHSNum	Mean of all years' number of high spells	$Year^{-1}$	Lawson et al. (2015)
Peak frequency	Frequency of peaks above three times the monthly median flow	Dimensionless	Pennino, McDonald, and Jaffe (2016)
RLF	Number of positive changes in flow from one day to the next	Dimensionless	Zhang et al. (2016)
RHF	Number of negative changes in flow from one day to the next	Dimensionless	Zhang et al. (2016)
NFLH	Number of negative and positive changes in flow from one day to the next	Dimensionless	Zhang et al. (2016)
R–B Index	Richard–Bakers Flashiness Index	Dimensionless	Cherkauer and Sinha (2010)
HP	High pulse count, frequency of events that exceed a threshold of two times mean annual flow	Not given but presumed dimensionless	Wenger et al. (2010)
Volume-to-peak ratio	Hydrograph volume divided by peak flow discharge	Dimensionless	Pennino et al. (2016)
Duration			
LowS75	Low-flow spell duration (<75th percentile)	Days	Zhang et al. (2016)
HigS25	High-flow spell duration (>25th percentile)	Days	Zhang et al. (2016)
ZeroN	Number of zero-flow days	Days	Zhang et al. (2016)
LFD	Low-flow duration	Weeks	Null et al. (2010)
Pdd	Dry-to-dry spell transition probability	Dimensionless	Wilby, Prudhomme, Parry, and Muchan (2015).
Pww	Wet-to-wet spell transition probability	Dimensionless	Wilby et al. (2015).
State			
Index of dryness	Potential evapotranspiration divided by precipitation	Dimensionless	Zhang et al. (2004)
Evapotranspiration ratio	Actual evapotranspiration divided by precipitation	Dimensionless	Zhang et al. (2004)
Evapotranspiration efficiency	Actual evapotranspiration divided by potential evapotranspiration	Dimensionless	Zhang et al. (2004)
Index of wetness	Precipitation divided by potential evapotranspiration	Dimensionless	Zhang et al. (2004)
Runoff ratio	Runoff total divided by precipitation total	Dimensionless	Mendoza et al. (2016), Ehsanzadeh et al. (2012), Ehsanzadeh, van der Kamp, and Spence (2016)
Rainfall elasticity	Fractional change in annual runoff divided by the fractional change in annual precipitation	Dimensionless	Chiew (2006), Vano, Das, and Lettenmaier (2012)
RAS	Relative aridity score	Dimensionless	Wade et al. (2013)
SPI	Standardized Precipitation Index	Dimensionless	McKee, Doesken, and Kleist (1993)
SMRI	Standardized Snow Melt and Rain Index	Dimensionless	(Staudinger, Stahl, & Seibert, 2014)
PDSI	Palmer Drought Severity Index	Dimensionless	Palmer (1965)
SWSI	Surface Water Supply Index	Dimensionless	Shafer and Dezman (1982)
SRI	Standardized Runoff Index	Dimensionless	Shukla and Wood (2008)
SPEI	Standardized Precipitation Evapotranspiration Index	Dimensionless	Vicente-Serrano, Begueria, and Lopez-Moreno (2010)

TABLE 1 (Continued)

Metric	Brief description	Unit	Reference
WBR	Normalized modified Thornthwaite water balance runoff	Dimensionless	Willmott, Rowe, and Mintz (1985), Dobrowski et al. (2013)
TM	Drought termination magnitude	%	Parry, Prudhomme, Wilby, and Wood (2016), Parry, Wilby, Prudhomme, and Wood (2016)
DM	Drought magnitude	%	
DDD	Drought development duration	Months	
DTD	Drought termination duration	Months	
Service and performance			
Floodplain performance threshold	Annual average of floodplain area that is inundated for at least *X* consecutive days (expressed as the magnitude of change relative to mean historical conditions)[c]	Dimensionless	Poff et al. (2016)
Magnitude of daily changes in outflows from a reservoir in response to upstream characteristics	Expressed as the magnitude of change relative to mean historical conditions	Dimensionless	Poff et al. (2016)
Mean annual water delivery to export service areas	Mean annual water delivery to export service areas	Not given	Brekke et al. (2009)
Mean end-of-“time-in-year”[d] upstream-of-storage location	As defined by name	Not given	Brekke et al. (2009)
Projected change in reservoir firm yield (difference between current and future yield)	As defined by name	Billions in cubic meters	Maldonado and Moglen (2013)
LoS	Level of service, for example, a target for the maximum annual probability of a shortage of given security.	Dimensionless	Hall et al. (2012).
DO	Deployable output—maximum rate that a system can supply water continuously through a dry period with a known or assumed severity	Can be defined as a volume or rate (unit not given)	Hall et al. (2012)
D_{DYA}	Theoretical dry year—dry year annual average unrestricted daily demand’	Rate (unit not given)	Hall et al. (2012)
Change in demand for water	For example, domestic demand for water	Liters per head per day	Wade et al. (2013)
H_A	Head-room allowance—difference between the water available for use (DO including raw-water inputs minus raw-water exports and outage) and the D_{DYA}	Volume (unit not given)	Hall et al. (2012)
WSCI	Water Supply Capacity Index—designed to assess ability to satisfy an estimated optimal water demand. Estimated from the ratio of water resources availability to water demand (which can be for a particular demand source, e.g., domestic demand)	Dimensionless	Collet, Ruelland, Estupina, Dezetter, and Servat (2015)
FUY	Frequency of unsatisfactory years, where unsatisfactory imply at least on occurrence of a WSCI below a high satisfaction rate	Dimensionless	Collet et al. (2015)
REL	Reliability—metric capturing how often a system fails	Dimensionless	Collet et al. (2015), Hashimoto, Stedinger, and Loucks (1982)
RES	Resilience—metric capturing how quickly a system returns to a satisfactory state once a failure has occurred	Dimensionless	Collet et al. (2015), Hashimoto et al. (1982)
VUL	Vulnerability—metric capturing how significant the likely consequences of failure may be	Dimensionless	Collet et al. (2015), Hashimoto et al. (1982)
RI	Robustness index—a metric that considers outcomes for performance metrics across a wide range of possible future climates	Dimensionless	Whateley, Steinschneider, and Brown (2014)
CRI	Same as RI, but probabilities are assigned to certain climate outcomes	Dimensionless	Whateley et al. (2014)

[a] Calculated for all seasons.

[b] Spring is used as an example here, but this metric was also calculated for other seasons.

[c] “X” refers to length of period considered in metric, in Poff et al. (2016) a value of 7 days was used.

[d] “Time-in-year” is inserted to indicate that this value is application dependent. In Brekke et al. (2009) the period is “September”.

peak flow per calendar year in time series from 4,262 hydrometric stations. The average timing of river flooding for each station was estimated as the average date for when floods have occurred in the observed records. By plotting the average timing for each station location using vector markers, geographical differences in seasonality of floods are easily identified; the angle and color of the vector denote when during the year flooding occurs, its length indicate the temporal distribution of floods during the year (figure 3).

2.2 | Magnitude

Magnitude metrics are simply measures that inform on the typical magnitude associated with specified thresholds (e.g., quantiles), or maximum and minimum levels. They are a means to quantify streamflow characteristics and can be used by, for example, water managers to detect potential trends and/or abrupt shifts in resources. For example, Harrigan, Murphy, Hall, Wilby, and Sweeney (2014) looked annual mean and high flows to understand how change points might be linked to combinations of climate and human-induced drivers and Kirono et al. (2014) used mean flow on a daily, annual, dry and wet season basis to support the formulation of climate adaptation policy for the Mamminasata metropolitan region of Indonesia. Awareness of shifts in these metrics can support planning activities around augmentation of resources or operations under a changing climate, but it is critical that the choice of flow metric reflects the intended application. For example, when seeking information about change to low-flow conditions, agricultural users may require information about average 7- or 30-day minimum flow, whereas municipal planners may be more concerned about a distribution-related magnitude, such as the flow occurring 25% of the time (Pal, Towler, & Livneh, 2015).

Magnitude metrics can reflect instantaneous or time-averaged discharge, reflecting peak behavior or the average flow over a day or multiple days. Often, these metrics are calculated from time series for multidecadal periods to avoid sensitivity to climate variability, but can also be calculated on an annual or seasonal basis to describe changes in, for example, deciles of daily flow (Rouge & Cai, 2014). Other examples include metrics derived from the flow duration curve, such as the flow duration curve median (FMM; estimated as the median of the flow duration curve in logarithmic space) intended to characterize mid-range flow behavior (Mendoza et al., 2016), or the proportion of flow in a specified month (e.g., Q_{Feb} of Booker and Woods (2014).

A range of magnitude-related metrics are provided in Table 1 (section Magnitude). Many of these are self-explanatory and need no further introduction in the text, such as percentile exceedances or percentile ranges. Often a combination of several metrics is used to quantify distributional characteristics. For example, Cherkauer and Sinha (2010) investigated climate change impacts on freshwater inflows to the Great Lakes in Midwestern United States and Canada using a suite of metrics including average annual or seasonal (3-month seasons) peak daily streamflow (peak flow); average annual or seasonal mean daily streamflow (mean flow); average annual or seasonal minimum daily stream flow (minimum flow); average annual or seasonal cumulative sum of flows above a specified daily flow level (at a level with 20% exceedance probability; high-flow sums); and average annual or seasonal cumulative sum of low-flow deficit (level set to 80% exceedance probability; low-flow deficit).

If interested in rare events rather than mean behavior, thresholds metrics can be selected that are relevant to the application at focus. For example, the fifth percentile level of normalized flow (Q05) was used to capture high-flows in a model evaluation for India (Pechlivanidis & Arheimer, 2015), the mean magnitude of high spells above the 95th percentile of the flow duration curve (HSPeak) was used to study the magnitude of flooding disturbance in a riparian zone in Australia. Focusing on low-level flows, the 95th percentile exceedance discharge (Q95) was used to describe periods with less water available for the environment in the UK Climate Change Act 2008 (Wade et al., 2013), and a 7-day mean annual low flow (Q_{MALF}) was used to define low-flow characteristics in a New Zealand-based hydrological modeling study by Booker and Woods (2014).

Metrics indicative of the state of a stream include the "flow duration curve low-segment volume" (FLV) and "fraction of time that daily streamflow exceeds mean streamflow for each year" (T_{Qmean}). The FLV was used by Mendoza et al. (2016) in Colorado, United States to represent long-term base flow, and the T_{Qmean} by Cherkauer and Sinha (2010) to denote redistributions of streamflow from base-flow to fast-response storm-flow.

2.3 | Extreme values

Extreme events can have significant hydrologic and economic impacts. Unsurprisingly, many metrics have been proposed to quantify such occasions (Table 1, section Extreme values). Typically, events with 10, 100, or even 500-year return periods (equivalent to 10%, 1%, and 0.2% annual exceedance probability) are cited. Common application areas for use of extreme metrics are reservoir and/or flood management, noting that reservoirs often have a dual function to mitigate flooding and secure water resources. Examples include the use of a 30-day 20-year low-flow event ($30Q_{20}$) when modeling of the

Occoquan reservoir in Virginia, United States (Maldonado & Moglen, 2013), and the maximum 3-day runoff volume during a 30-year period (3D30Y) as a proxy indicator for changes in flood risk. In Brekke et al. (2009), seasonal ratio changes of the metric 3D30Y between current-to-future time periods informed adjustments of flood control rules, reflecting change in seasons typically associated with flooding and refill. If focusing on storm-water infrastructure (drains, bridges, and so on) intensity–duration–frequency (IDF) curves are familiar engineering tools; their purpose being to characterize precipitation intensities of different storm durations and return periods (Tfwala, van Rensburg, Schall, Mosia, & Dlamini, 2017). However, the use of IDF curves depends on the stationarity assumptions inherent to extreme value theory; as such their adaptation for use in a climate change perspective is a subject of high interest (Cheng & AghaKouchak, 2014; Fadhel, Rico-Ramirez, & Han, 2017; Westra et al., 2014).

Estimating long return period events from an extreme value distribution fit to relatively short hydrological records assumes stationarity, hence there is considerable uncertainty (Schulz & Bernhardt, 2016). This may be manifested by poor estimates of the extreme value distribution parameters used to estimate the magnitude of required return period events. For instance, Tye and Cooley (2015) used a spatial extreme value model applied to rainfall time series across mountainous terrain in Colorado, United States. Despite this uncertainty, metrics can be used in a meaningful way for decision making. For example, changes in magnitude, or frequency of flooding can be connected to a geographic area and thus to infrastructure and economic impacts (Wobus et al., 2017).

Rouge and Cai (2014) used a "peak over threshold" (POT) method to identify temporal changes in the frequency of high-flow events in the Greater Chicago area, where "peaks" are defined as maximum streamflows on a centred 15-day window. The metric POT1 contains X peaks, where X is the number of years in the series, with an average of one peak per year. A metric POT2 then contains twice the numbers of peaks compared to POT1, and so on (the authors used POT1–5). The temporal distribution of the POT metrics can then be analyzed to study how high-flow events are distributed over the studied time horizon. In contrast, the use of POT in extreme value theory (e.g., Mailhot, Lachance-Cloutier, Talbot, & Favre, 2013) facilitates increased sample sizes compared with annual maxima, and hence parameter estimation, for distribution fitting. Its limitations include the influence of seasonality and temporal dependence within the data series, affecting bias and parameter estimation (e.g., Katz, Parlange, & Naveau, 2002).

Examination of seasonal variability in extremes can inform on linkages between flow and other seasonally dependent systems (such as ecosystems or resource allocation). For example, high-flows in winter and summer can impact fish populations as these may have a negative effect on fall-spawning and spring-spawning fish, respectively. Wenger et al. (2010) captured the frequency of high-flows using: the number of days in winter in the top first and fifth percentile annual flows (W99 and W95); the probability that a 1- or 2-year flow event occurs in winter (W1.5 and W2); and the frequency of summer flows above the annual fifth percentile (S95). [Note that, a 2-year flow event "W2" is the same as a 50th percentile event]. To reflect the low-flow characteristics that may limit fish population in summer, similar metrics to those for winter were calculated, namely: days with flows in summer less than the 10th and 20th percentile of the mean annual flow (S10, S20); or the 7-day average low-flow with a 10-year return period (7Q10).

2.4 | Variability

These metrics characterize temporal variability, attempting to capture day-to-day variability, flashiness or multidecadal climate-related variability (Table 1, section Variability). Straightforward day-to-day variability can be estimated by metrics such as the coefficient of variation (CV) of daily flow (CVDF) and low/high-flow spell count (LowC75/HigC25) (Zhang et al., 2016). Others are based on counts of the average annual or seasonal number of days with flows above/below the daily flow level (high-flow/low-flow count) (Cherkauer & Sinha, 2010). To study links between the hydrological regime and riparian ecosystem diversity in Australia, Lawson et al. (2015) described the average variability within mean daily flows for each season using the ratio of the CV of mean daily seasonal daily flow to the overall mean daily flow (CVMDFSpring), and variability in high flows by the mean annual frequency of high-spell periods (flow above 95th percentile) (MDFAnnHSNum), and its CV (CVAnnHSPeak) giving the interannual variability in high flows. Depending on the intent, variability metrics can take different forms. The next couple of paragraphs give a few examples.

To capture flashiness of flow magnitudes Mendoza et al. (2016) use a flow duration curve mid-segment slope (FMS) metric, defined as the ratio of the range of the logarithm of flows associated with the exceedance probability of 0.2 and 0.7 (e.g., $\log(Q_{0.2})$–$\log(Q_{0.7})$), and the exceedance probability range (e.g., 0.2–0.7). A similarly approach is used by Pechlivanidis and Arheimer (2015) and Viglione et al. (2013) in the use of their slope of the flow duration curve metric (m_{FDC}, Equation (2)).

$$m_{\mathrm{FDC}} = 100 \cdot \frac{Q_{30\%} - Q_{70\%}}{40 \cdot \overline{Q_{\mathrm{d}}}} \tag{2}$$

where $Q_{30\%}$ and $Q_{70\%}$ are, respectively, the daily flow magnitude that is exceeded 30% and 70% of the time, and $\overline{Q_{\mathrm{d}}}$ is the mean daily specific runoff. Interested in "flashy" high-flow events, Pennino et al. (2016) used metrics "peak frequency", estimated as the frequency of peaks above three times the monthly median flow, and the volume-to-peak ratio, defined as the hydrograph volume divided by the peak flow discharge. Cherkauer and Sinha (2010) used the Richards–Baker Flashiness Index (R–B Index), which is the ratio between the sum of the absolute values of day-to-day changes in daily discharge volumes and the total discharge volumes for each year or season. Wenger et al. (2010) applied the high pulse count (HP), defined as the frequency of events that exceed a threshold of twice mean annual flow. Reversals of flows can be captured by the number of positive/negative/total changes in flow from one day to the next (RLF/RHF/NFLH; i.e., frequency of reversals between higher and lower flows on consecutive days; Zhang et al., 2016).

Variability over multiannual and multidecadal timescales can be linked to large-scale ocean–atmosphere modes. Wilby and Quinn (2013) related annual maximum and peak over threshold series to an objective weather classification to evaluate long-term changes in concurrent multibasin flooding in the United Kingdom. Naturally, multidecadal variability is increasingly recognized as a key source of uncertainty in climate model projections (Deser, Knutti, Solomon, & Phillips, 2012; Deser, Phillips, Alexander, & Smoliak, 2014), sometimes overwhelming the differences within and between model ensembles (Deser, Phillips, Bourdette, & Teng, 2012; Sriver, Forest, & Keller, 2015). However, there are relatively few metrics to quantify this uncertainty due to the greater attention on hydrological variability over shorter-timescales, in addition to shortage of long historical records.

2.5 | Duration

Duration metrics specify the longevity or persistence of certain flow conditions (Table 1, section Duration). For instance, Zhang et al. (2016) used spell-length duration metrics (LowS75/HigS25; defined as days below the 75th and 25th percentile of flow, respectively) in combination with the number of zero-flow days (ZeroN). Null et al. (2010) examined the low-flow duration (LFD), a characteristic critical for water supply planning in montane ecosystems. Their LFD metric is the number of weeks (a minimum of three) with low-flow conditions (defined as weeks with <1% of the total discharge in that water year).

Wilby et al. (2015) analyzed the length of dry (below long-term average rainfall) and wet (above average rainfall) spells in the United Kingdom by quantifying observed dry-to-dry (Pdd) and wet-to-wet (Pww) season transitions. First, the frequency of dry-to-dry (or wet-to-wet) transitions is counted, then the Pdd (or Pww) metric is the fraction of dry-to-dry transitions relative to all transitions. These transition probabilities were used in a first-order Markov model to estimate the likelihood of very persistent dry-spells. For instance, the dry-spell duration with 100-year return period was found to be greater than 5.5 years. Such persistent events were observed in the 19th century but not yet in the modern era. The Pdd and Pww metrics can also be used as indicators of climate change.

2.6 | State

This class of metrics describes changes in the state of natural or managed systems using descriptions of water balance terms, moisture status, or aridity (Table 1, section State). As such, they are integrating conditions of some water balance components that might evolve over daily, seasonal, annual, or even multidecadal timescales. For instance, potential increases in evaporative losses under climate change are of significant concern for water resource management.

2.6.1 | Water balance terms

The long-term regional water balance can be characterized by the Budyko curve, a water balance framework describing the partitioning of rainfall into evapotranspiration as a function of an index of dryness (PET/rainfall). Zhang et al. (2004) used the water balance relationship underpinning the Budyko curve to assess impacts of catchment characteristics and climatic drivers on the partitioning of rainfall into evapotranspiration and runoff. They provide a summative view of the regional water balance using metrics such as the index of dryness, the evapotranspiration ratio (ET/rainfall), the evapotranspiration efficiency (ET/PET), and an index of wetness (rainfall/PET). When estimated over a multiyear period, such measures can be used to determine whether a catchment is water limited ($ET/p > 1$) or energy limited ($ET/p < 1$).

Other examples of water balance-type metrics are the runoff ratio (i.e., runoff depth divided by total precipitation depth) (Ehsanzadeh et al., 2012; Ehsanzadeh et al., 2016; Mendoza et al., 2016) and the rainfall elasticity (the ratio of the fractional change in annual runoff to the fractional change in annual precipitation) (Vano et al., 2012). These metrics can indicate sensitivity in runoff to a given change in rainfall, noting that such a change may also be influenced by surface-groundwater connectivity, land surface, and land-use change which are particularly likely from a long-term perspective (Chiew et al., 2014). The runoff ratio may also be normalized for changes in precipitation so that other hydrologic effects may be assessed despite a changing background-climate. For example, Biederman et al. (2015) used the runoff ratio to help quantify the effect of tree die-off on streamflow in western North America.

2.6.2 | Moisture status, aridity, or drought

The Relative Aridity Score (RAS) gives information about the warmth and dryness of a period relative to climatology (and thus is location specific). Wade et al. (2013) applied the RAS to gain insights into biophysical impacts on water resources as part of the 2013 UK Climate Change Risk Assessment (CCRA). Weights were assigned to the temperature and rainfall components of the RAS to reflect the higher importance given to rainfall in the UK water balance (Equation (3)):

$$\mathrm{RAS} = 0.4 \times \frac{T_{\mathrm{future}} - T_{61-90}}{\mathrm{SDT}_{61-90}} - 0.6 \times \frac{\mathrm{Rain}_{\mathrm{future}} - \mathrm{Rain}_{61-90}}{\mathrm{SD\,Rain}_{61-90}} \tag{3}$$

where T_{future} is the average annual temperature for the future period, $T_{61–90}$ and $\mathrm{SDT}_{61–90}$ are, respectively, the average and *SD* of temperature during the 1961–1990 baseline period. Similarly, $\mathrm{Rain}_{\mathrm{future}}$ denotes the average of annual totals for a future time period, $\mathrm{Rain}_{61–90}$ and SD $\mathrm{Rain}_{61–90}$ are the average and *SD* over the 1961–1990 period.

The RAS captures aridity rather than specific drought events and does not consider year-to-year variability. If hydrological drought is of interest, then other metrics cover the temporal and spatial extent, frequency, severity, and moisture depletion from upper soil levels, for example, as proposed by Mishra and Singh (2010; p. 207–209):

- The Standardized Precipitation Index (SPI) (McKee et al., 1993) which describes the probability of a specified precipitation amount within a specified period of time. A probability distribution is fitted to precipitation and transformed to a normal distribution with zero mean for the desired period. Negative (positive) values indicate less (more) precipitation than median precipitation. The main challenges concern obtaining a long precipitation record and finding an appropriate probability distribution with good fit to the data. The Standardized Melt and Rainfall Index (SMRI) (Staudinger et al., 2014) is a variation on the SPI that accounts for precipitation as snow, through including deficits in SM and rainfall to streamflow.
- The Palmer Drought Severity Index (PDSI; Palmer, 1965) is based on the water balance between soil moisture supply and demand in a two-layer soil model, given information about temperature and precipitation. Whilst used to identify regional drought, the PDSI is arguably more suited to agricultural drought applications. Moreover, it is assumed that all precipitation is in liquid form (which is problematic for cold and mountainous regions); runoff is supposed to occur when all soil layers are saturated; and the index can be slow to respond to developing and diminishing droughts. Furthermore, "severity" is considered to be equivalent across all regions. However, differences in agricultural practices, local water resource management practices and precipitation responses can all result in very different impacts for a given precipitation deficit (Towler & Lazrus, 2016).
- The Surface Water Supply Index (SWSI; Shafer & Dezman, 1982) is the monthly non-exceedance probability-based information about supply sources. The SWSI is mainly used to monitor changes in surface water supply for urban, industrial, or agricultural use. Due to high dependency on available information and, because factor weights assigned to sources may vary in time and space, comparisons across spatiotemporal scales are difficult.
- The Standardized Runoff Index (SRI; Shukla & Wood, 2008) is similar to the SPI, but estimated from streamflow time series and hence includes variability due to hydrologic processes.

Water balance metrics are sensitive to input data and formulation. For example, Abatzoglou, Barbero, Wolf, and Holden (2014) applied four drought metrics to 21 sites to evaluate whether their relationship with water-year streamflow depth varies across catchments in the Pacific North West of the United States. In addition to the PDSI and the SPI, the authors used the Standardized Precipitation Evapotranspiration Index (SPEI; which includes a simple water balance adjustment, where PET is removed from the precipitation; Vicente-Serrano et al., 2010), and a normalized modified Thornthwaite water balance runoff (WBR; Dobrowski et al., 2013; Willmott et al., 1985). Acknowledging some limitations in data, the authors found that drought indices including "atmospheric moisture demand" performed better than simpler indices, and PET estimated by Penman–Monteith lead to higher correlation between metrics and streamflow, particularly during the growing season.

The limitations of drought indices have been widely discussed. Trenberth et al. (2014) address the challenges of calculating and analyzing the PDSI at a global scale. Although they focus on estimation of PDSI using observational data, many issues are pertinent to climate change. For instance, the need to consider the exact formulation of the index (whether applicable/calibrated to the region of interest) and the method used to estimate ET (in terms of the different quality of constituent variables) (see e.g., Guo, Westra, & Maier, 2016). Vicente-Serrano, Van der Schrier, Beguería, Azorin-Molina, and Lopez-Moreno (2015) analyzed the sensitivity of drought indices to precipitation and reference ET and found low sensitivity of PDSI to the latter, whereas the SPEI showed equal sensitivity to both variables. Thus, care is needed when applying PDSI in regions where one may expect large changes to reference ET. Several authors noted that metrics based on ET estimated by the empirically based Thornthwaite method are very sensitive to temperature changes, leading to projections of unprecedented future drought (Feng, Trnka, Hayes, & Zhang, 2017; Sheffield, Wood, & Roderick, 2012).

A different type of limitation in drought analysis is the lack of understanding of the termination process. Parry, Prudhomme, et al. (2016) discussed the importance of understanding and characterizing this stage in a water management context. Stressing that drought termination is not simply an "end point," but rather a process of recovery to "normal" conditions in different sections of the hydrological system (e.g., groundwater, soil moisture, and streamflow). Several indices are proposed to define the drought development phase and its termination phase. These are based on monthly mean streamflow data, converted to percentage anomalies ($Z_{\%\text{anom i}}$) relative to a long-term average (LTA) (Parry, Prudhomme, et al., 2016; Parry, Wilby, et al., 2016):

- Termination magnitude (TM): the magnitude of $Z_{\%\text{anom i}}$ after a prespecified number of consecutive positive time steps (drought end reached).
- Drought magnitude (DM): the maximum negative $Z_{\%\text{anom i}}$, marks the end of the drought development phase.
- Drought development duration (DDD): the length of period between the start of the drought and the end of the drought development phase (maximum negative $Z_{\%\text{anom i}}$).
- Drought termination duration (DTD): the length of period between the start of the termination phase (the time step after maximum negative $Z_{\%\text{anom i}}$) and the end of the drought (after a prespecified number of consecutive positive $Z_{\%\text{anom i}}$).

2.7 | Service and performance

Metrics in this category focus on flow characteristics or measures that are relevant to operational matters and stakeholder interests (Table 1, section Service and Performance). Poff et al. (2016) give an example of applying metrics in an eco-engineering decision scaling (EEDS) framework for the Iowa River, where performance indicators were used to evaluate trade-offs between ecological functions and economic loss through flooding damage. Their first metric represents a simplified measure of floodplain function, set as the "historical annual average of floodplain area that is inundated for at least seven consecutive days". The second metric represents the magnitude of flow recession rates in reservoir outflow, which can have negative impacts on aquatic species. This was estimated as the magnitude of daily changes in outflows from the reservoir during periods when flows are released rapidly in response to upstream inflows. Co-evaluating these metrics allows water resource managers to allow for human needs whilst maintaining ecological function *despite* climate change. The subsections below give other examples of metrics supporting adaptation planning in terms of water supply reliability, water supply–demand, system sustainability, and performance.

2.7.1 | Water supply reliability

These metrics focus primarily on the supply side of water resources. For example, Brekke et al. (2009) used two metrics of water supply reliability to assess climate risks to reservoir operations in California, United States. These were the "mean annual water delivery to the export service areas," and the "mean end-of-September upstream-of-delta storage" in upstream reservoirs. When used together, they help to optimize current-year water delivery for drought protection (i.e., ensuring sufficient carry-over-storage for subsequent years), whilst not enhancing flood risk to downstream users. Maldonado and Moglen (2013) assessed climate and land-use change impacts on water resources in the Occoquan catchment, United States. They evaluated the projected change in reservoir firm yield—the difference between the historical and projected safe volume that can be drawn from a reservoir during different seasons and under various storage volumes.

Probabilistic metrics are also used to inform decisions in water resource management. For instance, the level of service (LoS) metric is defined as a "target for the maximum annual probability of a shortage of given security", such as annual probability of hosepipe bans not to exceed 0.05 (Hall et al., 2012, p. 120). Borgomeo et al. (2014) used LoS for water supply systems in London to assess frequencies of water shortages of varying severity with various portfolios of options. Block and Goddard (2012) applied a similar approach using precipitation exceedance probability curves to evaluate acceptable risk to reservoir operations and hydropower production in Ethiopia.

2.7.2 | Water supply and demand

To assess sustainability of water resources systems, operators need to consider supply and the ability of a system to meet different levels of demand. In England and Wales, water companies have a 25-year planning horizon (Hall et al., 2012). When supply is limited, water restrictions are put in place to conserve supply. Several metrics are used to summarize information about available resources and risks, such as deployable output (DO), defined as the maximum rate that a system can supply water continuously through a dry period with a known, or assumed, severity. The theoretical dry year (D_{DYA}) is defined as the "dry year annual average unrestricted daily demand". This conservative metric combines demographic and climate change impacts to assess possible onset of deficits in water security. Both metrics were used to assess risk to water availability under the UK Climate Change Risk Assessment (CCRA) (Wade et al., 2013). The CCRA also considered the measure "change in demand for water", defined as the domestic demand for water in liters per head per day.

To provide flexibility due to uncertainty in estimates, water resource management plans can include a head-room allowance (H_A) defined as the "difference between the water available for use (the DO including raw-water inputs minus raw-water exports and outage) and the D_{DYA}" (Hall et al., 2012). A target head-room may be estimated deterministically by assigning an acceptable probability (or level of risk) for exceeding H_A.

2.7.3 | System sustainability and performance

These metrics give insights into the robustness of a system. For example, Collet et al. (2015) used metrics of water supply and water sustainability for domestic, agricultural, and environmental stakeholders in the Hérault River catchment, France. The Water Supply Capacity Index (WSCI) assesses ability to satisfy an estimated water demand, defined as the ratio of water resources availability to water demand (which can be for a specified demand source, e.g., domestic use). Sustainability was assessed using "frequency of unsatisfactory years" (FUY) together with reliability, resilience, and vulnerability (RRV) metrics. The FUY is an indicator with a value in the range 0 to 5, and for a time series of Y years, is estimated as:

$$\mathrm{FUY} = \frac{5}{Y} \cdot \sum_{y=1}^{Y} \mathrm{UY}(y) \tag{4}$$

where UY(y) is 1 if the yth year is unsatisfactory else UY(y) is 0; where unsatisfactory implies at least one occurrence of a WSCI below a high satisfaction rate (e.g., meeting water demands without restriction). The metric "reliability" (REL) is a representation of the success of a system as a proportion of time spent in an unsatisfactory state with a value ranging from 0 to 1:

$$\mathrm{REL} = 1 - \frac{\sum_{j=1}^{M} d(j)}{T} \tag{5}$$

where M is the number of unsatisfactory periods and $d(j)$ is the length of the jth unsatisfactory period and T is the total length of the period. The resilience metric (RES, a value ranging from 0 to 1) is a ratio of number of unsatisfactory events relative to the duration of unsatisfactory periods, and is an indicator of how quickly on average a system recovers to a satisfactory state:

$$\mathrm{RES} = \frac{M}{\sum_{j=1}^{M} d(j)} \tag{6}$$

Finally, "vulnerability" (VUL) quantifies the severity of an unsatisfactory state. In Collet et al. (2015), VUL was given by the maximum difference between the complete satisfaction of estimated water demand and the actual water supply capacity in a year.

$$\mathrm{VUL} = max\left\{\sum\nolimits_{j \in d(j)} C(j) - X(j)\right\} \tag{7}$$

where C represents the water supply objective and X, the actual system performance during jth unsatisfactory period.

Many alternative formulations of RRVs have been used for water resource system performance evaluation since Hashimoto et al. (1982) demonstrated their application. For example, Goharian, Burian, Bardsley, and Strong (2016) extended the vulnerability metric, noting that basing it on severity alone can lead to incorrect quantification of system vulnerability. Rather, it could be a function of "exposure," "severity," and "potential severity". "Exposure" can be system failure due to climate change and "potential severity" a situation when an action leads to failure at a later time step. For example, if a reservoir is full, water might be released or bypassed, so unavailable at a future time when the same reservoir is at a critically low level (Goharian et al., 2016, p.6). Fundamentally, RRVs measure different aspects of water resource system performance, and together they are used to maximize performance outcomes under uncertainty due to different states of the world (Asefa, Clayton, Adams, & Anderson, 2014; Sandoval-Solis, McKinney, & Loucks, 2011).

Whateley et al. (2014) propose the robustness index (RI) as an alternative to the more traditional RRVs, suggesting that the latter are hampered by assumed stationarity of the hydroclimate system. Having identified a threshold for acceptable performance (e.g., water supply) model simulations explored system sensitivities within a plausible range of climate change. A more complex version of the index allows the RI to be weighted by various climate projections, thus taking into the account the probability of different climate futures (climate-informed RI, CRI).

Other robustness metrics are described and tested by Giuliani and Castelletti (2016) on the Lake Cosmo water reservoir operator. They compared the maximin and the maximax metric, the optimism–pessimism rule, the minimax regret metric, and the principal of insufficient reason. Their specifics and formulation are not explained here as these are generic maximization metrics rather than hydroclimate metrics. However, we note that these metrics "… can lead to different and mutually contradicting decisions" (Giuliani & Castelletti, 2016: p. 411).

3 | VALIDITY OF METRICS IN A CLIMATE CHANGE CONTEXT

Confidence in projections of metrics depends on ability to accurately simulate hydrological processes. Can the model fundamentally simulate the characteristic of interest? This is not a trivial task, for example, Pal et al. (2015) note that much is still unknown about fundamental hydrological processes affecting low-flows. Many regions have insufficient monitoring to fully understand, and thus represent the connectivity between surfacewater and groundwater, and the impact of human interaction through reservoir management and irrigation on water resources. Other challenges arise when extrapolating parameters to ungauged catchments (e.g., Hannaford, Holmes, Laizé, Marsh, & Young, 2013) or due to model sensitivity to nonstationarity in input data. For example, performance shortcomings are reported when models are calibrated to hydrological regimes different to those to which they are applied (Chiew et al., 2014; Milly et al., 2008; Milly et al., 2015; Vaze et al., 2010; Wilby, 2005). Furthermore, Mendoza et al. (2016) used different hydrological model structures and calibration techniques to assess impacts on annual water balance and other flow metrics. They showed that models with similar skill could lead to very different projections, highlighting the need to consider parameter uncertainty when estimating climate change impacts on water resources. Broderick et al. (2016) reached similar conclusions based on a multimodel and multicatchment study. They demonstrated the need to test the transferability of parameter sets between contrasting climate conditions and catchment types. They also used a multimodel ensemble in combination with an objective ensemble averaging technique to obtain robust estimates of future flow under a changing climate.

Whilst acknowledging the importance of "hydrological modeling challenges," we focus on issues that can violate the validity of a metric simply because of its use in a climate change context. Specifically, we consider validity from the perspective of formulation, dependency on climate input data, and the decision-making context.

3.1 | Formulation

Nonstationarity in the climate affects the estimation of metrics even for the historical period. Metrics describing the typical behavior of the hydrological regime such as average/high-/low-flows or rainfall-runoff ratio, require multidecadal time series for their estimation, while extreme value statistics require even longer periods. In a climate change context, it is questionable whether such metrics are robust due to expected trends in the data over such a time-period. Using shorter periods for estimation is possible, but then robustness is likely compromised by under-representation of climate variability. Hence, sensitivity testing of the metric to the length of the time-frame and trend effects is recommended. Pal et al. (2015) note that some low-flow metrics may not adequately capture the temporal shift in risk that is important to users. For example, a low-flow metric initially designed to regulate stream pollution is now applied as a generic low-flow measure, however current societal or environmental risks associated with low flow may be entirely different and should therefore have a formulation that is relevant to the current risk (Pal et al., 2015). It is also suggested that alternative methods may be sought from extreme value theory that allow for an assessment of the trend component in the estimation of very rare events (see e.g. van Haren, van Oldenborgh, Lenderink, and Hazeleger (2013).

Metrics may also be sensitive to season and water year definitions. Dery et al. (2009) note the sensitivity of simple timing metrics to the calendar definition of the water year (i.e., rain falling on the "wrong side" of the temporal divide could lead to misinterpretation of results). In addition, shifts in seasonality of weather patterns or hydrologic signals, such as earlier or slower snowmelt (Musselman, Clark, Liu, Ikeda, & Rasmussen, 2017), could alter interpretation of calendar dependent metrics. For example, peak flows in a month may become dependent more on rainfall than snowmelt.

If GCMs do not simulate the present seasonal cycle of temperature and rainfall accurately, this may reduce confidence in model projections (Moise et al., 2015) and hence derived metrics, such as snowmelt timing. Options may include adjusting the calendar window to account for seasonal bias or unusually early/late season onsets or use of less stringent definitions (e.g., winter half-year instead of a standard 3-month winter season).

3.2 | Climate input data

For many hydrological modeling frameworks, the input variables are total precipitation and PET. Whilst the former is a direct output from climate models (though may require summing up output in more than one variable, e.g., output from parameterized processes, such as convection schemes, and rainfall resolved directly by the microphysics scheme) estimates of PET equivalent to that expected by the hydrological model may need to be estimated. Note that whilst PET or evaporation is sometimes outputted, this may not correspond directly to what is expected for hydrological modeling. Furthermore, different models output differently, hence if using raw output, it is crucial to pay close attention to whether a variable is accumulated (if so over what time period) or instantaneously outputted. Also, check that the units of variables outputted by the climate models agree what is expected by the PET formulation. Finally, whilst the climate model can output on very high temporal resolution, the model may have less skill at the daily and sub-daily resolution compared to monthly resolution.

Ideally, all relevant (or at least first-order importance) physical processes would be realistic in GCM output. Often this is not the case. For example, the timing of high-flow events (e.g., JMaxF) may be strongly influenced by onset of SM in spring (freshets) or intense rainfall events. Local mountain snowpack melt and intense rainfall are two phenomena that require process representation beyond that feasible in GCMs (Kendon et al., 2014; Magnusson, Jonas, Lopez-Moreno, & Lehning, 2010). This is because the accumulation and wasting of the snowpack is influenced by meteorological (radiation, wind, precipitation), hydrological (melt-water), and soil (soil temperature and moisture) variables. Whilst statistical downscaling can improve resolution, maintaining intervariable relationships as well as spatial dependencies across multiple variables is not straightforward. Outputs from RCMs may offer the physical consistency needed (across variables and at the native resolution of the RCM), but typically RCM output requires bias correction (a process whereby the grid cell distribution of the simulated variable is statistically adjusted to better match historical observations). Such post-processing can violate such dependencies (Ehret, Zehe, Wulfmeyer, Warrach-Sagi, & Liebert, 2012; Gutmann et al., 2014; Maraun, 2013; Pierce, Cayan, Maurer, Abatzoglou, & Hegewisch, 2015) and recent work has demonstrated that the process itself can introduce implausible climate change signals (Maraun et al., 2017).

We stress that more precise information provided by downscaling does not equate to more accurate information about *change* in relevant water balance terms (e.g., Gutmann et al., 2012). Finer detail can be added by more-or-less complex methods and some capture only a few aspects of the GCM-simulated change (such as the mean change in a variable). More complex methods attempt to represent a fuller regional climatic response (Gutmann, Barstad, Clark, Arnold, & Rasmussen, 2016; Mearns et al., 2013; Prein et al., 2013), but may transfer model-specific biases (Ekström, Grose, & Whetton, 2015). The resulting hydrological metrics will directly depend on the capability of the downscaling techniques to capture change in relevant physical processes.

If the metric seeks to characterize a particular event, it is relevant to consider whether the climate model has realism at that event scale. For example, high-flow or extreme rainfall metrics are particularly sensitive to downscaling method as many systems generating intense rainfall are not well resolved by climate models. Fowler and Ekström (2009) found that regional climate models with 50 km spatial resolution did reasonably well in capturing UK winter extremes compared to summer extremes. This is due to the greater spatial footprint of the systems generating winter extremes (frontal passages) relative to summer extremes (convective storms). Models operating at much finer spatial resolutions (<10 km) can improve the representation of rainfall, particularly in mountainous terrain and at short temporal resolutions (Chan et al., 2013; Prein et al., 2013; Prein et al., 2015). However, convection resolving models (<2 km) have also been reported to simulate too intense convective rainfall (Kendon et al., 2014). A consequence of moving to finer resolved model output is the introduction of greater spatial variability in the output field. If the simulated rainfall events have spatial- or magnitude-related biases, this can lead to occasionally large biases in subsequent streamflow estimates (Kay, Rudd, Davies, Kendon, & Jones, 2015; Mass, Ovens, Westrick, & Colle, 2002).

Another aspect of model skill is the ability to represent wet-day occurrence. Potter and Chiew (2011) point to the importance of this characteristic because of its relevance to soil moisture and thus ability to estimate accurately flow in dry catchments. RCMs typically generate too much light rain (i.e., the drizzle effect; Kjellstrom et al., 2010), so post-processing may be required. Advances in modeling at finer resolutions appear to have resolved this limitation (Kendon, Roberts, Senior, & Roberts, 2012), but errors in heavier events may be introduced through overestimation of rainfall in convective storms. A common work-around is to assume that rain days are daily totals with magnitudes above a selected threshold. Kjellstrom et al. (2010) noted that low intensity rainfall did not significantly add to the daily total and used a threshold of 1 mm day^{-1} for their assessment of RCM statistics across Europe. However, the definition of the wet-day threshold determines the amount of data available for the analysis (Moberg & Jones, 2005; Zhang et al., 2011) and can affect extreme value estimates of peak over threshold totals (Schär et al., 2016).

Just as the number of wet-days are important, so is the sequencing of rain days or events (e.g., 7-day min, RLF, RHF, and NFLH). We note that metrics assuming a realistic representation of sequencing cannot be estimated using climate change information based on scaling of observed data (i.e., where observed data is scaled by a climate change factors). This is because the sequencing of events is unchanged to that of the observed. Distributional methods (involving different scaling -factors for different percentiles) may suffice for metrics capturing changes in frequency (e.g., high- or low-flow counts or fraction of time exceeding mean streamflow), but do not alter the sequencing. Even so, other properties such as the trend may be influenced by the binning structure applied (Michaels, Knappenberger, Frauenfeld, & Davis, 2004), and different users may regard different sections of the distribution to be of greater or lesser relevance.

The ability to reproduce the frequencies and range of natural climate variability is an important skill sought in climate models. This is particularly relevant for aridity and drought metrics influenced by oscillations in large-scale ocean–atmosphere systems such as the ENSO. Hope et al. (2015) report that, whilst GCMs were able to reproduce multiyear dry periods in southeast Australia, the length of these events is underestimated compared to observations covering the last 110 years. Even where GCMs have skill in representing such low-frequency variability, there is no guarantee that changes in future aridity metrics will be reliably derived.

3.3 | Decision making context

Metrics describing services and system performance can depend on the transferability of the current system to the future. For example, defining metrics such as DO for historic droughts can be problematic due to assumed stationarity in catchment response, or short records of regional climate variability (Hall et al., 2012). In terms of water resources, estimates are also needed about factors that influence demand. Turner et al. (2014) estimated a multidecadal water demand forecast for Melbourne, Australia, by accounting for trends in demographics and housing, industrial and commercial uses, water conservation technology uptake and leakage. Similar validity-concerns exist for operational rules. For example, Wade et al. (2013) note that environmental flow indices are based on historical flow regimes. However, the amount of water required to protect freshwater ecosystems against climate change may need to be adjusted for new policies, changing pressures and evolving aquatic communities. In the case of multi-objective schemes (such as a reservoir with water supply and flood defense roles), it is likely that operating rules will adjust to evolving priorities and/or climate conditions.

Future decision-making processes will reflect changing "values, rules, and knowledge" (vrk) (Gorddard, Colloff, Wise, Ware, & Dunlop, 2016). The vrk perspective emphasizes that rules-in-use are conditioned by current values and knowledge. Understanding the context of decision making enables stakeholders to influence agency and legitimacy of decision making. Such broad level research of the societal context can be combined with sensitivity testing of the system itself using perturbed climate input to explore system functions under a range of climate conditions (Guo, Westra, & Maier, 2018). The testing need not initially be constrained by the range of change projected by climate models but rather conducted to improve the knowledge of system behavior, particularly in data-poor regions—along similar sentiments of the first steps of analysis in the scenario neutral approach (Prudhomme, Wilby, Crooks, Kay, & Reynard, 2010).

4 | CONCLUSIONS

Hydroclimate metrics enable (a) monitoring of system performance, (b) evaluation of resources availability, and (c) assessment of model skill. Naturally, water managers are interested in how system-dependent metrics may evolve under future climate change. However, confidence in metrics depends on the skill of models to simulate hydrologic processes in a transient environment (Mendoza et al., 2016; Vaze et al., 2010) and on regional climate downscaling techniques to accurately project change at scales relevant to the intended application (Ekström, Grose, Heady, Turner, & Teng, 2016). When projecting a metric for future decades, uncertainty expands as more modeling frameworks are linked in the change estimation (e.g., emissions scenarios to climate system modeling).

Some assert that, given large uncertainties in projected climate change, adaptation decision-making should instead be grounded in knowledge of systems' behavior under different climate conditions. Such knowledge is gained through sensitivity (or stress) testing, and can support decisions on how to trade off optimal solutions with risk of increasing sub-optimality and the risk of infeasibility (Watkins & McKinney, 1997). These approaches may be supported by RRV, RI, and CRI metrics (table 7) that focus on robustness of decisions under uncertain futures (e.g., Weaver et al., 2013). Adopting a "graceful failure" approach, whereby the inevitable weakness in a system are acknowledged as part of the design (Tye, Holland, & Done, 2015) helps develop robustness to uncertainty. Nevertheless, driving hydrological models with climate change projections can yield useful information about the range of uncertainty and direction of change. Furthermore, such applications can contribute to improved understanding of key physical processes and their representation in models.

A unifying theme of this review is the need to evaluate the physical realism of the model information used in hydroclimatic metrics. Other challenges exist in the problem formulation, such as assumptions about future system performance, operating rules, levels of service, trade-offs, or reliability criteria. For metrics to give meaningful information to support decision making, the following general guidance is given with respect to development of hydrological metrics from climate change information:

- Establish the purpose and co-produce metrics with stakeholders to reflect their information needs and the specific decision-making contexts (whilst helping the same stakeholders to understand scientific limitations).
- Determine whether the metric is robust to transient climate change. Many metrics are meant to reflect long-term behavior and are estimated using multidecadal time series. That is, when using climate model output, consider presence of trend in data and consequential impacts on relevant metrics. Conduct sensitivity testing on the length of time-period used to estimate the metric and consider representation of natural climate variability.
- Determine whether climate models and post-processing methods provide valid information about changes in the characteristics of climate that the metric seeks to capture. For example, if the metric characterizes intense rainfall events, assess whether climate model outputs reproduce observed extremes under current climate—a necessary but insufficient test. Then, check the ability of the model/method to reflect change in the weather systems that generate extreme rainfall.
- When computing changes in metrics, always compare future values to the values computed in current climate with the same modeling setup; do not compare directly to observations of current climate.
- Understand that there is much scope for model structure uncertainty in hydroclimate metrics. Compare model output to observations to make assessment of model performance, compare outputs from different models to analyze structural sensitivity, and communicate this uncertainty clearly to stakeholders. When there are significant differences between model and observations in current climate, consider whether this metric is sufficiently well-simulated for the intended purpose.
- Review the validity of the problem formulation in a future world. Consider the decision-making context in which the metric is used. Contemplate whether rules and values that frame the decision-making process are likely to remain or change in a future world. Reflect on whether the values of future communities might change about environmental flows, or development of governance of water use.

To facilitate good practice in the use of metrics in climate change research, and to quantify the evolving uncertainty, researchers need to draw on the perspectives of hydrology, climate, and statistics. Such cross-disciplinary expertise can be contained in tools that enable consistent derivation of performance metrics such as HydroTest (Dawson, Abrahart, & See, 2007), or the R software package Evapotranspiration that calculates PET and AET using different formulations to facilitate estimation uncertainty (Guo et al., 2016). We further call for greater transparency and availability of long-term monitoring data held by water companies and utilities to support research on regional change to water supply and demand and, like Lettenmaier (2017), place priority on supporting existing and new hydrologic monitoring infrastructure. This review highlights the urgency to improve methods and models used to simulate hydrological systems under climate change; to include and reliably simulate all relevant relationships and factors. Such improvements could be facilitated by well-designed benchmarking studies underpinned by multimodel intercomparison work.

ACKNOWLEDGMENTS

Marie Ekström and Dewi Kirono were funded under the Earth Systems and Climate Change Hub, project 2.7 Refining Australia's water futures, of the Australian Government's National Environmental Science Programme. Marie Ekström is currently funded by the School of Earth and Ocean Sciences of Cardiff University and Cardiff University Water Research Institute. The National Center for Atmospheric Research is sponsored by the National Science Foundation. Mari Tye was supported by NSF EaSM3 Grant No. 1419563.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLES

An appraisal of downscaling methods used in climate change research
Climate projections for ecologists
Added value in regional climate modeling

REFERENCES

Abatzoglou, J. T., Barbero, R., Wolf, J. W., & Holden, Z. A. (2014). Tracking interannual streamflow variability with drought indices in the U.S. Pacific northwest. *Journal of Hydrometeorology*, *15*(5), 1900–1912. https://doi.org/10.1175/jhm-d-13-0167.1

Asefa, T., Clayton, J., Adams, A., & Anderson, D. (2014). Performance evaluation of a water resources system under varying climatic conditions: Reliability, resilience, vulnerability and beyond. *Journal of Hydrology*, *508*, 53–65. https://doi.org/10.1016/j.jhydrol.2013.10.043

Block, P., & Goddard, L. (2012). Statistical and dynamical predictions to guide water resources in Ethiopia. *Journal of Water Resources Planning and Management*, *138*(3), 287–298. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000181

Biederman, J. A., Somor, A. J., Harpold, A. A., Gutmann, E. D., Breshears, D. D., Troch, P. A., … Brooks, P. D. (2015). Recent tree die-off has little effect on streamflow in contrast to expected increases from historical studies. *Water Resources Research*, *51*(12), 9775–9789. https://doi.org/10.1002/2015WR017401

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., … Živković, N. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Booker, D. J., & Woods, R. A. (2014). Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. *Journal of Hydrology*, *508*, 227–239. https://doi.org/10.1016/j.jhydrol.2013.11.007

Borgomeo, E., Hall, J. W., Fung, F., Watts, G., Colquhoun, K., & Lambert, C. (2014). Risk-based water resources planning: Incorporating probabilistic nonstationary climate uncertainties. *Water Resources Research*, *50*(8), 6850–6873. https://doi.org/10.1002/2014WR015558

Brekke, L. D., Maurer, E. P., Anderson, J. D., Dettinger, M. D., Townsley, E. S., Harrison, A., & Pruitt, T. (2009). Assessing reservoir operations risk under climate change. *Water Resources Research*, *45*, 16 p. https://doi.org/10.1029/2008wr006941

Broderick, C., Matthews, T., Wilby, R. L., Bastola, S., & Murphy, C. (2016). Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resources Research*, *52*(10), 8343–8373. https://doi.org/10.1002/2016WR018850

Chan, S. C., Kendon, E. J., Fowler, H. J., Blenkinsop, S., Ferro, C. A. T., & Stephenson, D. B. (2013). Does increasing the spatial resolution of a regional climate model improve the simulated daily precipitation? *Climate Dynamics*, *41*(5–6), 1475–1495. https://doi.org/10.1007/s00382-012-1568-9

Cheng, L., & AghaKouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports*, *4*, 7093. https://doi.org/10.1038/srep07093

Cherkauer, K. A., & Sinha, T. (2010). Hydrologic impacts of projected future climate change in the Lake Michigan region. *Journal of Great Lakes Research*, *36*, 33–50. https://doi.org/10.1016/j.jglr.2009.11.012

Chiew, F. H. S. (2006). Estimation of rainfall elasticity of streamflow in Australia. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *51*, 613–625.

Chiew, F. H. S., Kirono, D. G. C., Kent, D. M., Frost, A. J., Charles, S. P., Timbal, B., … Fu, G. (2010). Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates. *Journal of Hydrology*, *387*(1–2), 10–23. https://doi.org/10.1016/j.jhydrol.2010.03.025

Chiew, F. H. S., Potter, N. J., Vaze, J., Petheram, C., Zhang, L., Teng, J., & Post, D. A. (2014). Observed hydrologic non-stationarity in far south-eastern Australia: Implications for modelling and prediction. *Stochastic Environmental Research and Risk Assessment*, *28*(1), 3–15. https://doi.org/10.1007/s00477-013-0755-5

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., … Brekke, L. D. (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, *2*(2), 55–64. https://doi.org/10.1007/s40641-016-0034-x

Clow, D. W. (2010). Changes in the timing of snowmelt and streamflow in Colorado: A response to recent warming. *Journal of Climate*, *23*(9), 2293–2306. https://doi.org/10.1175/2009jcli2951.1

Collet, L., Ruelland, D., Estupina, V. B., Dezetter, A., & Servat, E. (2015). Water supply sustainability and adaptation strategies under anthropogenic and climatic changes of a meso-scale Mediterranean catchment. *Science of the Total Environment*, *536*, 589–602. https://doi.org/10.1016/j.scitotenv.2015.07.093

Court, A. (1962). Measures of streamflow timing. *Journal of Geophysical Research*, *67*(11), 4335. https://doi.org/10.1029/JZ067i011p04335

Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software*, *22*(7), 1034–1052. https://doi.org/10.1016/j.envsoft.2006.06.008

Dery, S. J., Stahl, K., Moore, R. D., Whitfield, P. H., Menounos, B., & Burford, J. E. (2009). Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. *Water Resources Research*, *45*, 11 p. https://doi.org/10.1029/2008wr006975

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, *2*(11), 775–779. https://doi.org/10.1038/nclimate1562

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3–4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Deser, C., Phillips, A. S., Alexander, M. A., & Smoliak, B. V. (2014). Projecting north American climate over the next 50 years: Uncertainty due to internal variability. *Journal of Climate*, *27*(6), 2271–2296. https://doi.org/10.1175/jcli-d-13-00451.1

Dobrowski, S. Z., Abatzoglou, J., Swanson, A. K., Greenberg, J. A., Mynsberge, A. R., Holden, Z. A., & Schwartz, M. K. (2013). The climate velocity of the contiguous United States during the 20th century. *Global Change Biology*, *19*(1), 241–251. https://doi.org/10.1111/gcb.12026

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS opinions: "should we apply bias correction to global and regional climate model data?". *Hydrology and Earth System Sciences*, *16*(9), 3391–3404. https://doi.org/10.5194/hess-16-3391-2012

Ehsanzadeh, E., van der Kamp, G., & Spence, C. (2012). The impact of climatic variability and change in the hydroclimatology of Lake Winnipeg watershed. *Hydrological Processes*, *26*(18), 2802–2813. https://doi.org/10.1002/hyp.8327

Ehsanzadeh, E., van der Kamp, G., & Spence, C. (2016). On the changes in long-term streamflow regimes in the North American Prairies. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *61*(1), 64–78. https://doi.org/10.1080/02626667.2014.967249

Ekström, M., Grose, M., Heady, C., Turner, S. W. D., & Teng, J. (2016). The method of producing climate change datasets impacts the resulting policy guidance and chance of mal-adaptation. *Climate Services*, *4*, 13–29. https://doi.org/10.1016/j.cliser.2016.09.003

Ekström, M., Grose, M. R., & Whetton, P. H. (2015). An appraisal of downscaling methods used in climate change research. *Wiley Interdisciplinary Reviews: Climate Change*, *6*, 301–319. https://doi.org/10.1002/wcc.339

Fadhel, S., Rico-Ramirez, M. A., & Han, D. (2017). Uncertainty of intensity–duration–frequency (IDF) curves due to varied climate baseline periods. *Journal of Hydrology*, *547*, 600–612. https://doi.org/10.1016/j.jhydrol.2017.02.013

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

Feng, S., Trnka, M., Hayes, M., & Zhang, Y. (2017). Why do different drought indices show distinct future drought risk outcomes in the U.S. Great Plains? *Journal of Climate*, *30*(1), 265–278. https://doi.org/10.1175/jcli-d-15-0590.1

Fowler, H. J., & Ekström, M. (2009). Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *International Journal of Climatology*, *29*(3), 385–416. https://doi.org/10.1002/joc.1827

Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L., & Peterson, T. J. (2016). Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. *Water Resources Research*, *52*(3), 1820–1846. https://doi.org/10.1002/2015WR018068

Giuliani, M., & Castelletti, A. (2016). Is robustness really robust? How different definitions of robustness impact decision-making under climate change. *Climatic Change*, *135*(3–4), 409–424. https://doi.org/10.1007/s10584-015-1586-9

Goharian, E., Burian, S. J., Bardsley, T., & Strong, C. (2016). Incorporating potential severity into vulnerability assessment of water supply systems under climate change conditions. *Journal of Water Resources Planning and Management*, *142*(2), 04015051. https://doi.org/10.1061/(asce)wr.1943-5452.0000579

Gorddard, R., Colloff, M. J., Wise, R. M., Ware, D., & Dunlop, M. (2016). Values, rules and knowledge: Adaptation as change in the decision context. *Environmental Science & Policy*, *57*, 60–69. https://doi.org/10.1016/j.envsci.2015.12.004

Guo, D., Westra, S., & Maier, H. R. (2018). An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies. *Journal of Hydrology*, *556*, 887–890. https://doi.org/10.1016/j.jhydrol.2016.03.025

Guo, D., Westra, S., & Maier, H. R. (2016). An R package for modelling actual, potential and reference evapotranspiration. *Environmental Modelling & Software*, *78*, 216–224. https://doi.org/10.1016/j.envsoft.2015.12.019

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gutmann, E., Barstad, I., Clark, M., Arnold, J., & Rasmussen, R. (2016). The intermediate complexity atmospheric research model (ICAR). *Journal of Hydrometeorology*, *17*(3), 957–973. https://doi.org/10.1175/jhm-d-15-0155.1

Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., & Rasmussen, R. M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*, *50*(9), 7167–7186. https://doi.org/10.1002/2014WR015559

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Gochis, D. J., Clark, M. P., … Thompson, G. (2012). A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *Journal of Climate*, *25*(1), 262–281. https://doi.org/10.1175/2011jcli4109.1

Hall, J. W., Watts, G., Keil, M., de Vial, L., Street, R., Conlan, K., … Kilsby, C. G. (2012). Towards risk-based water resources planning in England and Wales under a changing climate. *Water and Environment Journal*, *26*(1), 118–129. https://doi.org/10.1111/j.1747-6593.2011.00271.x

Hannaford, J., Holmes, M. G. R., Laizé, C. L. R., Marsh, T. J., & Young, A. R. (2013). Evaluating hydrometric networks for prediction in ungauged basins: A new methodology and its application to England and Wales. *Hydrology Research*, *44*(3), 401–418. https://doi.org/10.2166/nh.2012.115

Harrigan, S., Murphy, C., Hall, J., Wilby, R. L., & Sweeney, J. (2014). Attribution of detected changes in streamflow using multiple working hypotheses. *Hydrology and Earth System Sciences*, *18*(5), 1935–1952. https://doi.org/10.5194/hess-18-1935-2014

Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982). Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, *18*(1), 14–20. https://doi.org/10.1029/WR018i001p00014

Hidalgo, H. G., Das, T., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Barnett, T. P., … Nozawa, T. (2009). Detection and attribution of streamflow timing changes to climate change in the Western United States. *Journal of Climate*, *22*(13), 3838–3855. https://doi.org/10.1175/2009JCLI2470.1

Hope, P., Timbal, B., Hendon, H. Ekstrom, M., Moran, R., Manton, M., Lucas, C., Ngyen, H., Lim, E.P., uo, J.-J., Liu, G., Zhao, M., Fiddes, S., Kirono, D., Wilson, L., Potter, N., Teng, J.,. (2015). *Victorian Climate Initiative annual report 2014–2015.* Retrieved from http://www.cawcr.gov.au/projects/vicci/wp-content/uploads/2015/11/BRR-005.pdf

Hurd, B., & Rouhi-Rad, M. (2013). Estimating economic effects of changes in climate and water availability. *Climatic Change*, *117*(3), 575–584. https://doi.org/10.1007/s10584-012-0636-9

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

Kay, A. L., Rudd, A. C., Davies, H. N., Kendon, E. J., & Jones, R. G. (2015). Use of very high resolution climate model data for hydrological modelling: Baseline performance and future flood changes. *Climatic Change*, *133*(2), 193–208. https://doi.org/10.1007/s10584-015-1455-6

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., & Senior, C. A. (2014). Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, *4*(7), 570–576. https://doi.org/10.1038/nclimate2258

Kendon, E. J., Roberts, N. M., Senior, C. A., & Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, *25*(17), 5791–5806. https://doi.org/10.1175/jcli-d-11-00562.1

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D., & Marsh, N. (2010). Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. *River Research and Applications*, *26*(2), 137–156. https://doi.org/10.1002/rra.1249

Kirono, D. G. C., Larson, S., Tjandraatmadja, G., Leitch, A., Neumann, L., Maheepala, S., … Selintung, M. (2014). Adapting to climate change through urban water management: A participatory case study in Indonesia. *Regional Environmental Change*, *14*(1), 355–367. https://doi.org/10.1007/s10113-013-0498-3

Kjellstrom, E., Boberg, F., Castro, M., Christensen, J., Nikulin, G., & Sanchez, E. (2010). Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Research*, *44*(2–3), 135–150. https://doi.org/10.3354/cr00932

Lawson, J. R., Fryirs, K. A., Lenz, T., & Leishman, M. R. (2015). Heterogeneous flows foster heterogeneous assemblages: Relationships between functional diversity and hydrological heterogeneity in riparian plant communities. *Freshwater Biology*, *60*(11), 2208–2225. https://doi.org/10.1111/fwb.12649

Lettenmaier, D. P. (2017). Observational breakthroughs lead the way to improved hydrological predictions. *Water Resources Research*, *23*, 1100, 2591–2597. https://doi.org/10.1002/2017WR020896

Magnusson, J., Jonas, T., Lopez-Moreno, I., & Lehning, M. (2010). Snow cover response to climate change in a high alpine and half-glacierized basin in Switzerland. *Hydrology Research*, *41*(3–4), 230–240. https://doi.org/10.2166/nh.2010.115

Mailhot, A., Lachance-Cloutier, S., Talbot, G., & Favre, A.-C. (2013). Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach. *Journal of Hydrology*, *476*, 188–199. https://doi.org/10.1016/j.jhydrol.2012.10.036

Maldonado, P. P., & Moglen, G. E. (2013). Low-flow variations in source water supply for the Occoquan reservoir system based on a 100-year climate forecast. *Journal of Hydrologic Engineering*, *18*(7), 787–796. https://doi.org/10.1061/(asce)he.1943-5584.0000623

Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, *26*(6), 2137–2143. https://doi.org/10.1175/jcli-d-12-00821.1

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., … Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*, 764–773. https://doi.org/10.1038/nclimate3418

Mass, C. F., Ovens, D., Westrick, K., & Colle, B. A. (2002). Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, *83*(3), 407–430. https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales.* Paper presented at the Eighth Conference on Applied Climatology, Anaheim, CA, US.

Mearns, L. O., Sain, S., Leung, L. R., Bukovsky, M. S., McGinnis, S., Biner, S., … Sloan, L. (2013). Climate change projections of the North American regional climate change assessment program (NARCCAP). *Climatic Change*, *120*(4), 965–975. https://doi.org/10.1007/s10584-013-0831-3

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*, *30*(7), 1071–1095. https://doi.org/10.1002/hyp.10684

Michaels, P. J., Knappenberger, P. C., Frauenfeld, O. W., & Davis, R. E. (2004). Trends in precipitation on the wettest days of the year across the contiguous USA. *International Journal of Climatology*, *24*(15), 1873–1882. https://doi.org/10.1002/joc.1102

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., … Krysanova, V. (2015). On critiques of "Stationarity is dead: Whither water management?". *Water Resources Research*, *51*(9), 7785–7789. https://doi.org/10.1002/2015WR017408

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, *391*(1–2), 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Moberg, A., & Jones, P. D. (2005). Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99. *International Journal of Climatology*, *25*(9), 1149–1171. https://doi.org/10.1002/joc.1163

Moise, A., Wilson, L., Grose, M., Whetton, P., Watterson, I., Bhend, J., … Rafter, T. (2015). Evaluation of CMIP3 and CMIP5 models over the Australian region to inform confidence in projections. *Australian Meteorological and Oceanographic Journal*, *65*(1), 19–53.

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*, 214–219. https://doi.org/10.1038/nclimate3225

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., … Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–2500. https://doi.org/10.1175/jhm-d-15-0026.1

Null, S. E., Viers, J. H., & Mount, J. F. (2010). Hydrologic response and watershed sensitivity to climate warming in California's Sierra Nevada. *PLoS One*, *5*(3), e9932. https://doi.org/10.1371/journal.pone.0009932

Pal, I., Towler, E., & Livneh, B. (2015). How Can We Better Understand Low River Flows as Climate Changes? EOS, 96, Published on August 6, 2015. doi:https://doi.org/10.1029/2015EO033875

Palmer, W. C. (1965). Meteorological drought.

Parry, S., Prudhomme, C., Wilby, R. L., & Wood, P. J. (2016). Drought termination: Concept and characterisation. *Progress in Physical Geography*, *40*(6), 743–767. https://doi.org/10.1177/0309133316652801

Parry, S., Wilby, R. L., Prudhomme, C., & Wood, P. J. (2016). A systematic assessment of drought termination in the United Kingdom. *Hydrology and Earth System Sciences*, *20*(10), 4265–4281. https://doi.org/10.5194/hess-20-4265-2016

Pechlivanidis, I. G., & Arheimer, B. (2015). Large-scale hydrological modelling by using modified PUB recommendations: The India-HYPE case. *Hydrology and Earth System Sciences*, *19*(11), 4559–4579. https://doi.org/10.5194/hess-19-4559-2015

Pennino, M. J., McDonald, R. I., & Jaffe, P. R. (2016). Watershed-scale impacts of stormwater green infrastructure on hydrology, nutrient fluxes, and combined sewer overflows in the mid-Atlantic region. *Science of the Total Environment*, *565*, 1044–1053. https://doi.org/10.1016/j.scitotenv.2016.05.101

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, *16*(6), 2421–2442. https://doi.org/10.1175/jhm-d-14-0236.1

Poff, N. L., Brown, C. M., Grantham, T., Matthews, J. H., Palmer, M. A., Spence, C. M., … Baeza, A. (2016). Sustainable water management under future uncertainty with eco-engineering decision scaling. *Nature Climate Change*, *6*, 25–34. https://doi.org/10.1038/nclimate2765

Potter, N. J., & Chiew, F. H. S. (2011). An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models. *Water Resources Research*, *47*(12), 11 p. https://doi.org/10.1029/2010wr010333

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., & Georgievski, G. (2013). Added value of convection permitting seasonal simulations. *Climate Dynamics*, *41*(9–10), 2655–2677. https://doi.org/10.1007/s00382-013-1744-6

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., … Leung, R. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, *53*(2), 323–361. https://doi.org/10.1002/2014RG000475

Prudhomme, C., Wilby, R., Crooks, S., Kay, A., & Reynard, N. (2010). Scenario-neutral approach to climate change impact studies: Application to flood risk. *Journal of Hydrology*, *390*(3–4), 198–209. https://doi.org/10.1016/j.jhydrol.2010.06.043

Reed, D. W. (1999). *Flood estimation handbook (procedures for flood frequency estimation): Overview*. Wallingford, England: Institute of Hydrology.

Rouge, C., & Cai, X. M. (2014). Crossing-scale hydrological impacts of urbanization and climate variability in the greater Chicago area. *Journal of Hydrology*, *517*, 13–27. https://doi.org/10.1016/j.jhydrol.2014.05.005

Sandoval-Solis, S., McKinney, D. C., & Loucks, D. P. (2011). Sustainability index for water resources planning and management. *Journal of Water Resources Planning and Management-Asce*, *137*(5), 381–390. https://doi.org/10.1061/(asce)wr.1943-5452.0000134

Schär, C., Ban, N., Fischer, E. M., Rajczak, J., Schmidli, J., Frei, C., … Zwiers, F. W. (2016). Percentile indices for assessing changes in heavy precipitation events. *Climatic Change*, *137*(1), 201–216. https://doi.org/10.1007/s10584-016-1669-2

Schulz, K., & Bernhardt, M. (2016). The end of trend estimation for extreme floods under climate change? *Hydrological Processes*, *30*(11), 1804–1808. https://doi.org/10.1002/hyp.10816

Schuster, Z. T., Potter, K. W., & Liebl, D. S. (2012). Assessing the effects of climate change on precipitation and flood damage in Wisconsin. *Journal of Hydrologic Engineering*, *17*(8), 888–894. https://doi.org/10.1061/(asce)he.1943-5584.0000513

Shafer, B. A., & Dezman, L. E. (1982). Development of a Surface Water Supply Index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. Paper presented at the 50th Annual Western Snow Conference, Reno, NV.

Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, *491*(7424), 435–438. http://www.nature.com/nature/journal/v491/n7424/abs/nature11575.html#supplementary-information

Shukla, S., & Wood, A. W. (2008). Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical Research Letters*, *35*(2), 7 p. https://doi.org/10.1029/2007gl032487

Sriver, R. L., Forest, C. E., & Keller, K. (2015). Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophysical Research Letters*, *42*(13), 5468–5476. https://doi.org/10.1002/2015GL064546

Staudinger, M., Stahl, K., & Seibert, J. (2014). A drought index accounting for snow. *Water Resources Research*, *50*(10), 7861–7872. https://doi.org/10.1002/2013WR015143

Teng, J., Vaze, J., Chiew, F. H. S., Wang, B., & Perraud, J.-M. (2012). Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff. *Journal of Hydrometeorology*, *13*(1), 122–139. https://doi.org/10.1175/jhm-d-11-058.1

Tfwala, C. M., van Rensburg, L. D., Schall, R., Mosia, S. M., & Dlamini, P. (2017). Precipitation intensity-duration-frequency curves and their uncertainties for Ghaap plateau. *Climate Risk Management*, *16*, 1–9. https://doi.org/10.1016/j.crm.2017.04.004

Towler, E., & Lazrus, H. (2016). Increasing the usability of drought information for risk management in the Arbuckle Simpson Aquifer, Oklahoma. *Climate Risk Management*, *13*, 64–75. https://doi.org/10.1016/j.crm.2016.06.003

Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, *4*(1), 17–22. https://doi.org/10.1038/nclimate2067

Turner, S. W. D., Marlow, D., Ekstroem, M., Rhodes, B. G., Kularathna, U., & Jeffrey, P. J. (2014). Linking climate projections to performance: A yield- based decision scaling assessment of a large urban water resources system. *Water Resources Research*, *50*(4), 3553–3567. https://doi.org/10.1002/2013WR015156

Tye, M. R., & Cooley, D. (2015). A spatial model to examine rainfall extremes in Colorado's front range. *Journal of Hydrology*, *530*, 15–23. https://doi.org/10.1016/j.jhydrol.2015.09.023

Tye, M. R., Holland, G. J., & Done, J. M. (2015). Rethinking failure: Time for closer engineer–scientist collaborations on design. *Proceedings of the Institution of Civil Engineers - Forensic Engineering*, *168*(2), 49–57. https://doi.org/10.1680/feng.14.00004

USGS. (2016). Explanations for the National Water Conditions. Retrieved from http://water.usgs.gov/nwc/explain_data.html

van Haren, R., van Oldenborgh, G. J., Lenderink, G., & Hazeleger, W. (2013). Evaluation of modeled changes in extreme precipitation in Europe and the Rhine basin. *Environmental Research Letters*, *8*(1), 7 p. https://doi.org/10.1088/1748-9326/8/1/014053

Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, *13*(3), 932–949. https://doi.org/10.1175/JHM-D-11-069.1

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., & Teng, J. (2010). Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, *394*(3–4), 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018

Vicente-Serrano, S. M., Begueria, S., & Lopez-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009jcli2909.1

Vicente-Serrano, S. M., Van der Schrier, G., Beguería, S., Azorin-Molina, C., & Lopez-Moreno, J.-I. (2015). Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *Journal of Hydrology*, *526*, 42–54. https://doi.org/10.1016/j.jhydrol.2014.11.025

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, *17*(6), 2263–2279. https://doi.org/10.5194/hess-17-2263-2013

von der Ohe, P. C., Prub, A., Schafer, R. B., Liess, M., de Deckere, E., & Brack, W. (2007). Water quality indices across Europe - a comparison of the good ecological status of five river basins. *Journal of Environmental Monitoring*, *9*(9), 970–978. https://doi.org/10.1039/b704699p

Wade, S. D., Rance, J., & Reynard, N. (2013). The UK climate change risk assessment 2012: Assessing the impacts on water resources to inform policy makers. *Water Resources Management*, *27*(4), 1085–1109. https://doi.org/10.1007/s11269-012-0205-z

Watkins, D. W., & McKinney, D. C. (1997). Finding robust solutions to water resources problems. *Journal of Water Resources Planning and Management-Asce*, *123*(1), 49–58. https://doi.org/10.1061/(asce)0733-9496(1997)123:1(49)

Weaver, C. P., Lempert, R. J., Brown, C., Hall, J. A., Revell, D., & Sarewitz, D. (2013). Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(1), 39–60. https://doi.org/10.1002/wcc.202

Wenger, S. J., Luce, C. H., Hamlet, A. F., Isaak, D. J., & Neville, H. M. (2010). Macroscale hydrologic modeling of ecologically relevant flow metrics. *Water Resources Research*, *46*, 10 p. https://doi.org/10.1029/2009wr008839

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., … Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. https://doi.org/10.1002/2014rg000464

Whateley, S., Steinschneider, S., & Brown, C. (2014). A climate change range-based method for estimating robustness for water resources supply. *Water Resources Research*, *50*(11), 8944–8961. https://doi.org/10.1002/2014wr015956

Wilby, R. L. (2005). Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes*, *19*(16), 3201–3219. https://doi.org/10.1002/hyp.5819

Wilby, R. L. (2010). Evaluating climate model outputs for hydrological applications. *Hydrological Sciences Journal*, *55*(7), 1090–1093. https://doi.org/10.1080/02626667.2010.513212

Wilby, R. L., Prudhomme, C., Parry, S., & Muchan, K. G. L. (2015). Persistence of hydrometeorological droughts in the United Kingdom: A regional analysis of multi-season rainfall and river flow anomalies. *Journal of Extreme Events*, *02*(02), 1550006. https://doi.org/10.1142/s2345737615500062

Wilby, R. L., & Quinn, N. W. (2013). Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. *Journal of Hydrology*, *487*, 109–121. https://doi.org/10.1016/j.jhydrol.2013.02.038

Willmott, C. J., Rowe, C. M., & Mintz, Y. (1985). Climatology of the terrestrial seasonal water cycle. *Journal of Climatology*, *5*(6), 589–606.

Wobus, C., Gutmann, E. D., Jones, R. G., Rissing, M., Mizukami, N., Lorie, M., … Martinich, J. (2017). Modeled changes in 100 year flood risk and asset damages within mapped floodplains of the contiguous United States. *Natural Hazards and Earth System Sciences*, *17*(12), 2199–2211. https://doi.org/10.5194/nhess-17-2199-2017

Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., & Briggs, P. R. (2004). A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research*, *40*(2), 14 p. https://doi.org/10.1029/2003WR002710

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., … Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

Zhang, Y., Shao, Q., Zhang, S., Zhai, X., & She, D. (2016). Multi-metric calibration of hydrological model to capture overall flow regimes. *Journal of Hydrology*, *539*, 525–538. https://doi.org/10.1016/j.jhydrol.2016.05.053

How to cite this article: Ekström M, Gutmann ED, Wilby RL, Tye MR, Kirono DGC. Robustness of hydroclimate metrics for climate change impact research. *WIREs Water*. 2018;5:e1288. https://doi.org/10.1002/wat2.1288