npj | climate and atmospheric science

www.nature.com/npjclimatsci

ARTICLE OPEN

Check for updates

Leveraging global climate models to assess multi-year hydrologic drought

Michael J. F. Vieira[1]✉ and Tricia A. Stadnyk[2]

Global climate models (GCMs) offer value for assessments of future water supply and multi-year hydrologic drought. Leveraging GCM data, we develop and analyze global scenarios of mean annual runoff over a span of 640 years. Runoff data from eighteen GCMs are evaluated for skill and bias-adjusted to reflect observations. Unprecedented projections of mean runoff, drought severity, and drought duration are found for 37%, 28%, and 23% of analyzed global land area, respectively, with regions on all continents presenting a risk of a drier future. Conversely, northern latitudes show evidence of increasing runoff, less severe, and shorter-duration droughts. Outside these regions, projections are either indistinguishable from internal climate variability or unreliable due to conflicting signal-to-noise ratios and ensemble agreement. Our analysis contributes to a global gap in understanding future multi-year hydrologic droughts, which can pose significant socio-economic risks.

npj Climate and Atmospheric Science (2023)6:179; https://doi.org/10.1038/s41612-023-00496-y

INTRODUCTION

Understanding water supply reliability and the occurrence of drought is important from social and economic perspectives. Droughts are often categorized as meteorological, agricultural, hydrological[1], and socio-economic when moisture deficits affect daily life or cause financial burdens. Impacts associated with all drought types can be devastating to human health[2,3] and broader global sustainable development goals[4]. From an economic perspective, prolonged drought can impact electricity prices[3,5,6] and cost billions of dollars[7–9].

Climate models are valuable tools to support assessments of future water supply[10,11] and various drought types[12–15]. Compared to earlier models, finer spatial resolutions[16], improved representation of land surface hydrology[17], and the availability of multi-century simulations[18] increase the appeal of Global Climate Models (GCMs) for hydrologic analyses. Although the opportunity for improvement remains[19–21], leveraging available GCM data can provide meaningful information, especially in data-sparse regions.

Future streamflow assessments typically use calibrated hydrologic models driven by climate model-derived temperature and precipitation for a region[22,23]. This approach incorporates fine hydrologic detail but can be computationally prohibitive at the global scale. Global Hydrologic Models (GHMs) and Land Surface Models (LSMs) facilitate analysis of larger domains[24–26] but may be subject to limited observations, exclusion of fine hydrologic detail, and less rigorous calibration[20]. GHMs have been used to project future changes in sub-annual hydrologic drought[27–29] without extension to multi-annual events. Beyond temperature and precipitation, climate models simulate runoff through physics-based representation of land surface processes which is important for modeling land-atmosphere energy exchange and freshwater discharge to oceans[13,30]. Direct use of climate model runoff has the benefit of preserving dynamically simulated atmosphere, ocean, and land surface feedbacks, which play an important role in climate development[31] and drought intensification[32].

Advances in land surface representation within climate models have spurred the direct use of climate model runoff as an alternative or complementary approach to traditional streamflow projection development[10,12,15,17,33–35]. While existing studies provide useful climate change analyses for sub-annual hydrologic patterns in specific regions over common time horizons (usually ≤30 years); there remains a gap in projecting multi-year hydrologic drought at the global scale.

We combine pre-industrial control (piControl), historic, and future GCM simulations to explore patterns of runoff change in 1460–2100. GCM skill is evaluated against pseudo-observed runoff products known as LORA[36] and GRUN[37] over 1980–2012, where LORA is also used for bias adjustment. Projected changes (signals; 2021–2100 versus 1941–2020) are compared against internal climate variability (noise; 1460–1940) and interpreted alongside ensemble agreement to demonstrate the pragmatic use of mean annual runoff scenarios to assess multi-year hydrologic drought. Runoff scenarios are analyzed globally (1° resolution) and within sixteen river basins (>600,000 km^2), chosen to provide wide coverage and include regions with notable water management infrastructure (Fig. 1). When future emissions scenarios RCP4.5 and RCP8.5 are viewed as equiprobable, unprecedented change in mean annual runoff (MAR), mean drought severity (SEV), and mean drought duration (DUR) is found for 37%, 28%, and 23% of global land area, respectively. Projections outside these regions are characterized as unreliable or indistinguishable from internal climate variability, which can still offer value for long-term planning.

RESULTS

Climate model skill

Evaluating skill informs practical applications of GCMs by providing insight regarding model abilities, shortcomings, and effectiveness of bias adjustment, and reminds users that models and observations are imperfect. Even among common variables like temperature and precipitation, GCMs exhibit errors in space and time[38–40] and we recognize that skill in one variable may correlate with skill in other variables[27,40–42]. Skill dependency may

[1]Water Resources Department, Manitoba Hydro, 360 Portage Ave (15), Winnipeg, MB R3C0G8, Canada. [2]Department of Geography, University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada. ✉email: mvieira@hydro.mb.ca

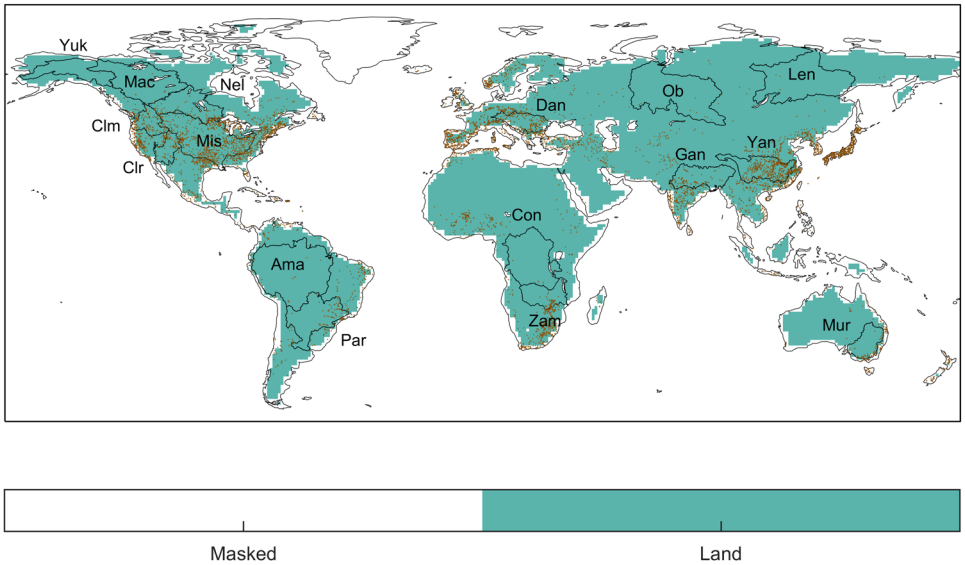

Fig. 1 Land mask. Sixteen large river basins overlain on a merged global land-sea mask (1° latitude by 1° longitude resolution), showing land grids that are consistent among all global climate models and the pseudo-observed datasets (see Methods). Brown points show 7250 dams from version 1.3 of the Global Reservoir and Dam Database (GRanD[101]).

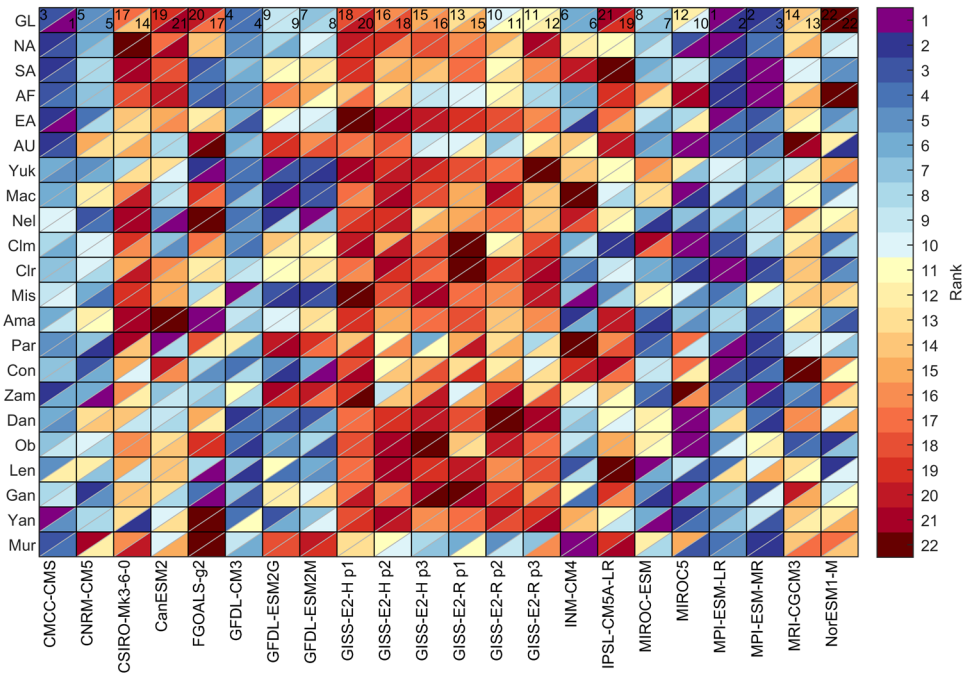

Fig. 2 Ranked skill by global climate model and region. GCM skill ranked by relative RMSE computed as the average of six runoff indices ($RMSE_{R\text{-}All}$; see Methods). Rows represent spatial domains as Global (GL), NA (North America), SA (South America), AF (Africa), EA (Eurasia), AU (Australia), and 16 river basins shown in Fig. 1. Columns represent the GCM simulations (prior to bias adjustment). Each cell is divided diagonally to show rank with respect to two pseudo-observed datasets (LORA[36] as the upper left portion of the cell and GRUN[37] as the lower right portion of the cell). A value of 1 indicates the best-performing model. Numeric ranks are shown for GL (top row) as a reference for colors in subsequent rows.

have an undesired effect if errors in one water balance component influence or compensate for errors in another component[21], but here we look exclusively at GCM skill in simulating runoff patterns.

Figure 2 ranks skill for eighteen GCMs (listed in Supplementary Table 1) based on relative RMSE ($RMSE_R$) in 22 regions for an aggregated measure of six indices (see Methods). Supplementary Figs. 1–4 and Supplementary Tables 2 and 3 present error magnitudes and demonstrate that bias adjustment successfully reduced error in most indices. Comparing results against an alternate pseudo-observed runoff dataset revealed only minor differences in global ranks. However, for some regions, models, and indices, Supplementary Table 2 reveals that GCMs can exhibit similar or better RMSE compared to an alternate pseudo-observed runoff product which is consistent with another study of GCM runoff skill[43].

At the global scale, MPI-ESM-LR, MPI-ESM-MR, and CMCC-CMS rank 1, 2, and 3 (respectively) and are among the best-performing GCMs across all six indices (Fig. 2; Supplementary Fig. 1). Top

performing GCMs at the global scale tend to perform well at the continental and river basin scales. Reasonable global skill can coincide with poor skill within more localized regions, however, and poor global skill can occasionally be influenced by poor performance in a specific region(s). In general, we find that RMSE and $RMSE_R$ metrics are, in some cases, sensitive to exceptionally poor performance in a small number of grids (e.g., see Supplementary Discussion).

RMSE is computed grid-by-grid and measures a GCM's ability to capture spatial patterns. Interpretation of RMSE is complemented by visual comparison of skill indices (maps; Supplementary Fig. 2), providing a spatially distributed view of GCM performance. GCMs broadly replicate wet and dry regions, but there are nuances at the individual grid scale, specifically, the size and location of wet and dry areas are imperfect. Interestingly, we find that NorESM1-M (ranked last) outperforms MPI-ESM-LR (ranked first) with respect to spatially averaged global mean, minimum, and maximum runoff (Supplementary Fig. 2), but has less skill in capturing global spatial patterns.

We demonstrate a range of runoff skills within the GCM ensemble and find that, overall, GCMs are capable of broadly capturing runoff patterns, including wet and dry regions. Our findings are generally consistent with GCM MAR skill evaluated at the continental scale[10]; however, it should be recognized that findings can be sensitive to selected skill indices, metrics, and reference datasets[43]. Through bias adjustment, GCM runoff errors in four of six annual runoff indices (mean, minimum, maximum, and coefficient of variation) are drastically improved, but errors in lag-1 autocorrelation and the average number of cumulative dry years remain (Supplementary Fig. 3, Supplementary Fig. 4, Supplementary Table 3). Reducing error through bias adjustment improves the utility of GCM runoff data for practical evaluation of hydrologic characteristics, but future work is needed to explore bias adjustment techniques that improve other skill indices relevant to drought analysis.

Continental hotspots with unprecedented projections

Figure 3 shows that future projections of MAR, coefficient of variability (CV), SEV, and DUR for most grids are either: indistinguishable from internal climate variability (most GCM signals lie within the noise) or unreliable (most GCM signals exceed noise but disagree on direction of change). While these findings can provide meaningful information for practitioners, we focus primarily on remaining regions that combine strong signals with good ensemble agreement on the direction of change and are herein classified as unprecedented. Unprecedented changes are common among MAR, SEV, and DUR but quite rare for CV. Regions with unprecedented change in one variable, however, do not always align with unprecedented change among other variables (e.g., increasing MAR may not align with decreasing SEV). Limited regions exhibiting unprecedented change underscore uncertainties and challenges in using GCMs for climate adaptation planning, but our analysis highlights regions where practitioners may benefit from greater confidence in projections.

Wetter conditions emerge for higher latitudes and portions of western and eastern North America, which may be a continuation of observed climatic changes such as glacier melt and increasing precipitation[14,44]. For water management facilities in northern Quebec (Fig. 1), gradual runoff increases (Supplementary Fig. 5 panel a) lead to reduced drought. Some southern U.S. States and portions of Mexico show evidence of a drier future consistent with the "drier in dry, wetter in wet" paradigm[45]. In these arid regions where natural variability already includes extreme drought events[8], gradual runoff reductions (Supplementary Fig. 5 panel b) coupled with increasing temperature could amplify impacts.

Isolated regions in South America are projected to dry, with some areas also projecting increased inter-annual runoff variability. Near Buenos Aires, there is evidence of increasing runoff, coupled with drought changes that are indistinguishable from internal climate variability. Gradual runoff declines in the Patagonia region (Supplementary Fig. 5 panel c) may result from reductions in precipitation[46] combined with changes in vegetation[47] and possibly exacerbated by continued depletion of ice fields[44].

Regions in North and South Africa are projected to dry, broadly following historic "dry gets drier" patterns[48]. Notably, these regions, which are particularly vulnerable to drought[9], overlap future hot spots for increased meteorological drought[14,49]. Projections for east-central Africa show increasing MAR, which deviates from some published drying patterns[48]; however, coinciding unreliable projections of drought make for a complex interpretation of future drought risk[15]. Projected wetting in this arid region may be, in part, attributed to vegetation changes resulting in runoff gains[47].

Two predominant patterns emerge within Eurasia: Northern regions are projected to become wetter, and west-central regions are projected to dry. This dry hotspot is consistent with projections of precipitation[16], sub-annual drought[9,15], and multi-year meteorological drought[49], which may exacerbate existing drought risk[50]. Looking jointly at MAR and SEV, we find additional northern regions projected to become wetter and an expanded west-central Eurasian region projected to become drier. Looking exclusively at unprecedented changes in MAR, the ensemble projects wetter conditions for some regions in southern Eurasia but with no associated unprecedented change in drought. Unprecedented signals in regions with dense populations and water management infrastructure pose opportunities and challenges depending on adaptive capacity. Drought risk challenges[9] may be especially relevant for humid regions previously thought to be less susceptible to the notion of dry getting drier[45,48].

Among continents, Australia has the smallest fraction of grids exhibiting unprecedented change. Projections for four hydrologic characteristics are predominantly within the range of noise due to internal climate variability, which may be a testament to the wide range of natural variability in Australian drought[6]. Runoff time series from individual GCMs reveal an ensemble dichotomy where some models simulate more stationary, linear patterns, and others show lower frequency signatures of high and low runoff (Supplementary Fig. 5 panel i). Multi-decadal periodicity creates challenges in bias adjustment to a relatively short (33-year) period and demonstrates the importance of interpreting climate change signals alongside internal climate variability.

Regional patterns and magnitudes of change found herein (e.g., wetting of northern latitudes, drying in central Eurasia) generally agree with earlier studies of annual runoff indices[11,35,51], sub-annual low flow indicators[27,28], and sub-annual drought indices[12,14,15]. Our findings confirm previously established projections and show that unprecedented areas of wetting and drying patterns generally extend to multi-year hydrologic drought events.

A practical perspective to leverage runoff and drought projections

For water resource planning in large watersheds, grid-scale changes may be less relevant than projections at the river basin scale. This is especially important for institutions responsible for water management infrastructure located on main stems at outlets of large rivers.

Figure 4 illustrates time series of bias-adjusted runoff in sixteen large river basins. In all basins, the ensemble median of mean runoff was relatively stationary up until 1941, with some trends emerging post-1941. Trends in projected mean runoff changes are most obvious for the Yukon, Mackenzie, Ob, Lena, and Ganges-Brahmaputra (increasing) and Danube (decreasing) river basins

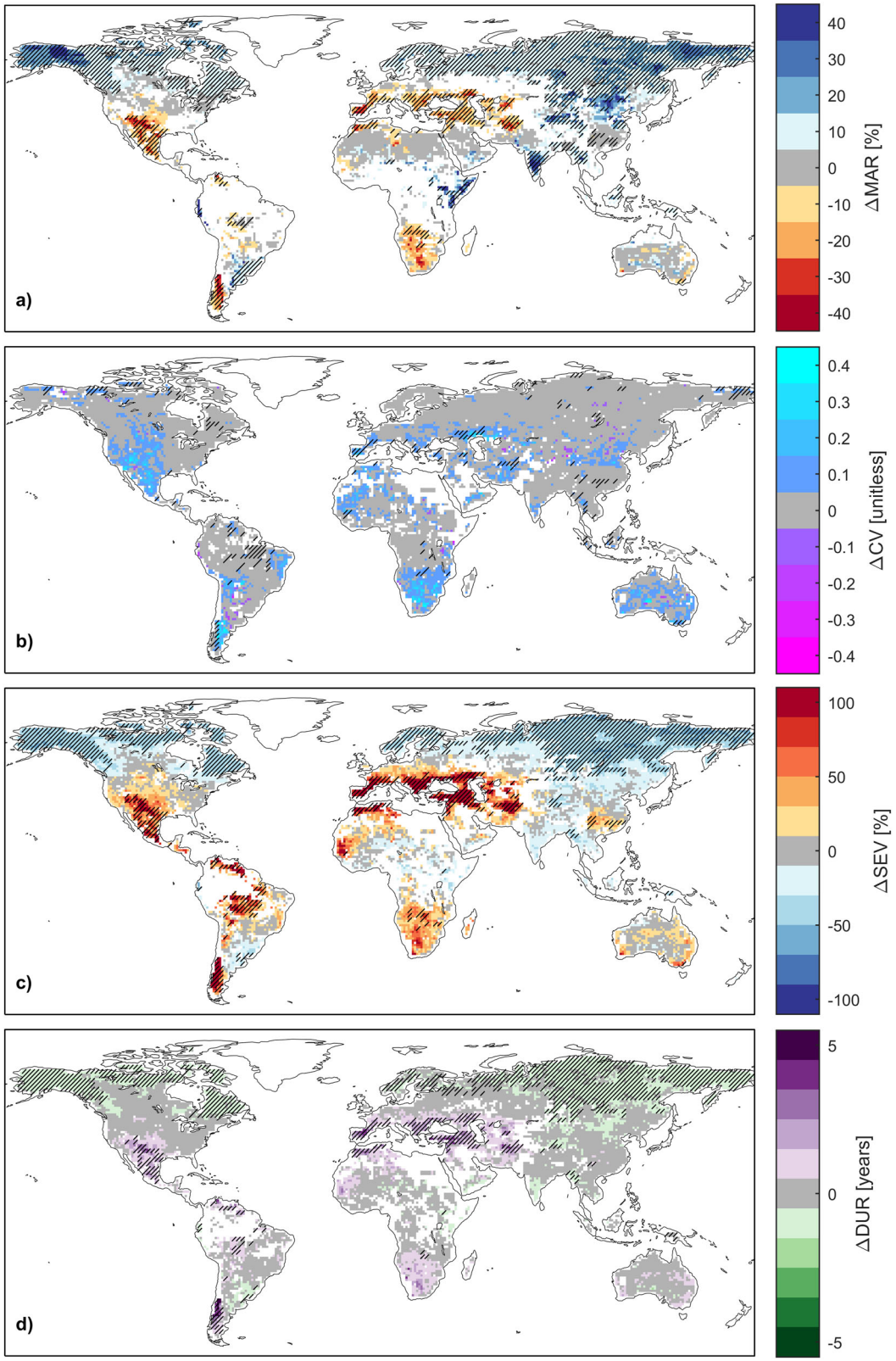

Fig. 3 Projected change in hydrologic characteristics. Ensemble median (RCP4.5 and RCP8.5) projected change (Δ) in mean annual runoff (MAR; panel **a**), coefficient of variation (CV; panel **b**), mean drought severity (SEV; panel **c**) and duration (DUR; panel **d**). Colors represent the magnitude of change and are shown for the 2021–2100 period relative to the 1941–2020 baseline. Hatching distinguishes regions with evidence of unprecedented change, white grids demark water, masked areas, or regions with unreliable projections, and other colored regions show areas where the projected change is indistinguishable from internal climate variability. See Methods for further information.

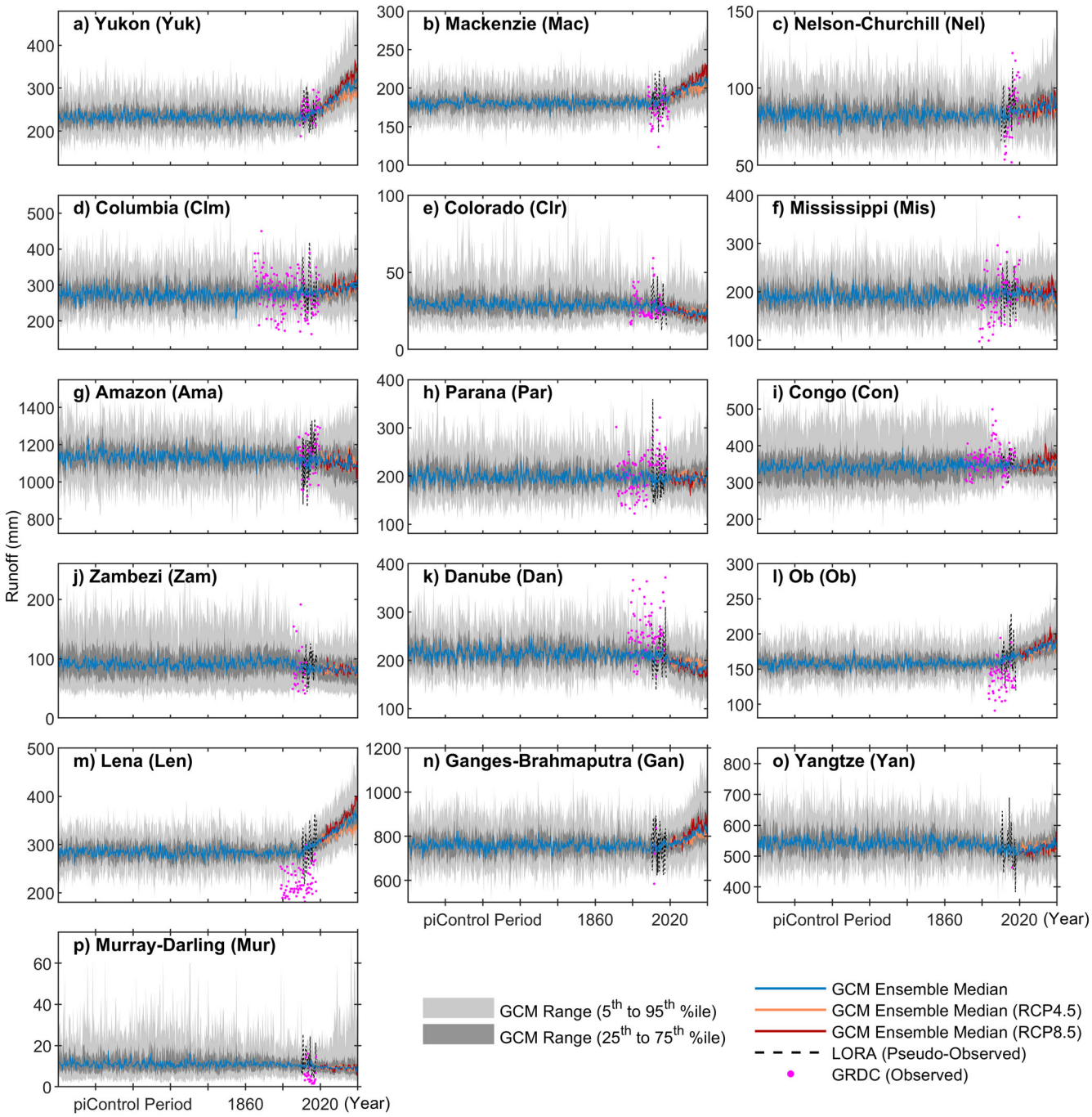

Fig. 4 Mean annual runoff time series. Time series of simulated runoff in sixteen river basins (shown in separate panels **a**–**p**). Light and dark gray bands show the 5th to 95th and 25th to 75th percentile ranges, respectively, from all bias-adjusted GCM simulations. The blue line shows the GCM ensemble median computed from all bias-adjusted simulations, while the orange and red lines represent the RCP4.5 and RCP8.5 ensemble median, respectively. The black dotted line shows the LORA (pseudo-observed) data which is complemented with GRDC mean annual runoff observations (streamflow divided by gauged drainage area). All panels share the same *x*-axis (time), but y-axes (mean annual runoff) are independent for each panel.

from 1941 to 2100. Patterns in the other nine basins are less (visually) apparent due to greater variability in the GCM ensemble regarding the direction of future change. Signal-to-noise ratios (SNRs) and ensemble agreement help distinguish results for individual basins (Table 1, Supplementary Table 4).

In all 16 river basins, projected changes in MAR were mostly found to be either unprecedented or unreliable. CV projections are mostly indistinguishable from internal climate variability, and projected changes in drought characteristics (SEV and DUR) are reported as either unprecedented, unreliable, or indistinguishable. Climate change adaptation decisions become more complex when projection information is unreliable or indistinguishable from internal variability, but identification of such regions with higher uncertainty can support cases for stress testing or further research. For example, in the presence of unreliable projections, practitioners may be inclined to undertake more detailed, finer-resolution hydrologic modeling studies to better grasp uncertainties. When projections are indistinguishable from internal

Table 1. Ensemble median projected changes (Δ; 2021–2100 vs. 1941–2020) for four hydrologic characteristics in sixteen large river basins.

		Yuk	Mac	Nel	Clm	Clr	Mis	Ama	Par	Con	Zam	Dan	Ob	Len	Gan	Yan	Mur
ΔMAR (%)	All[a]	21.4[b]	9.9[b]	5.9[b]	4.7[b]	−9.5[c]	−2.3[c]	−1.4[c]	2.6[c]	1.1[c]	−8.2[c]	−8.8[b]	8.5[b]	15.8[b]	5.4[b]	−0.2[c]	−8.9[c]
	RCP4.5	15.4[b]	7.5[b]	5.9[b]	3.4[b]	−9.5[c]	−2.3[c]	−0.9[c]	3.1[c]	−1.2[c]	−8.2[c]	−5.2[b]	7.9[b]	12.3[b]	3.7[b]	0.8[d]	−5.2[c]
	RCP8.5	26.6[b]	12.3[b]	5.3[c]	5[b]	−12[b]	−0.8[c]	−2.9[c]	1.5[c]	2.6[c]	−8.9[c]	−8.9[b]	10.2[b]	20.4[b]	8.8[b]	−2.3[c]	−11.1[c]
ΔCV *10[2e]	All[a]	2.1[b]	1.3[d]	1.4[d]	1.4[d]	4[d]	1.7[d]	0.4[d]	0.8[d]	1.3[b]	3.9[d]	1.9[b]	1.4[d]	0.8[d]	0.7[d]	1.7[d]	6.8[d]
	RCP4.5	0.9[d]	0.9[d]	1.4[d]	1.3[d]	4[d]	1.7[d]	0.3[d]	0.8[d]	0.3[d]	2.4[d]	1.1[d]	0.7[d]	0[d]	0.3[d]	1.7[d]	4.9[d]
	RCP8.5	4.5[b]	2.3[b]	1.5[d]	1.6[d]	3.8[d]	1.5[d]	0.5[d]	0.7[d]	3.2[b]	6.4[b]	2.7[b]	1.6[b]	1.8[d]	1.2[d]	1.5[d]	8.3[d]
ΔSEV (%)	All[a]	−59.3[b]	−45.6[b]	−25.5[d]	−22[d]	99.7[b]	15.1[c]	15.5[c]	−10.6[d]	8.3[c]	60.4[b]	86[b]	−38.7[b]	−60.1[b]	−31.7[d]	9.6[b]	53.2[c]
	RCP4.5	−59.4[b]	−45.6[b]	−25.5[d]	−25.2[d]	61[d]	15.1[d]	10.1[c]	−14.2[d]	10.6[c]	56[b]	64.9[b]	−38.7[b]	−58.5[b]	−25.2[d]	−0.5[d]	48.9[c]
	RCP8.5	−59.2[b]	−45.2[b]	−25.4[c]	−20.3[d]	131.2[b]	16.4[c]	59.6[c]	−4.6[d]	3.2[c]	66.3[b]	101.6[b]	−40.9[b]	−62.6[b]	−36.3[d]	24.2[b]	53.4[b]
ΔDUR (years)	All[a]	−1[b]	−0.7[b]	−0.5[d]	−0.6[d]	1.1[d]	0.2[d]	0.1[c]	−0.2[d]	−0.1[c]	0.8[b]	1.3[b]	−0.7[b]	−1[b]	−0.3[d]	0[d]	0.8[c]
	RCP4.5	−1[b]	−0.6[b]	−0.5[d]	−0.6[d]	0.7[d]	0.2[d]	0.1[d]	−0.3[d]	0.1[c]	0.7[b]	0.5[d]	−0.8[b]	−1[b]	−0.3[d]	0[d]	0.6[c]
	RCP8.5	−1[b]	−0.7[b]	−0.4[d]	−0.6[d]	1.9[b]	0.1[c]	0.2[c]	−0.1[d]	−0.2[c]	1[b]	1.5[b]	−0.6[b]	−1.1[b]	−0.5[d]	0[c]	0.9[b]

[a]Ensemble median projection for the entire ensemble, including both RCP4.5 and RCP8.5.
Projected change is characterized as unprecedented[b], or unreliable[c], indistinguishable[d].
[e]CV values are multiplied by 10^2 to reduce the display of decimal places.

variability, practitioners may be inclined to invest in a more comprehensive understanding of historical variabilities, such as through paleo-studies.

In some regions, careful decisions regarding climate adaptation may be supported by the presence of unprecedented projections. For example, water managers and planners with interests in the Yukon, Mackenzie, Ob, and Lena river basins may find assurance in projected declines in multi-year hydrologic drought-related risk. Water managers and planners operating near the outlets of the Zambezi and Danube rivers may similarly take measures to address increasing risk[9] associated with water availability and drought, such as the development of early detection, real-time monitoring, and forecasting initiatives. We find that projected changes for larger areas are dampened compared to individual grids, which may result from wet grids compensating for dry grids or be an artifact of performing bias adjustment independently for each grid. This feature is illustrated in Fig. 3, which shows ensemble median projections exceeding a 3-year change in DUR, and SEV increases on the order of 100%, meaning that the average future drought could be twice as severe as the average drought defined by the 1941–2020 period. However, projections at the river basin scale (Table 1; Supplementary Table 4) are less alarming. In many cases, we also find that a future with greater atmospheric greenhouse gas concentrations (i.e., RCP8.5) corresponds to higher magnitude changes that are unprecedented over broader regions (Supplementary Table 6; Supplementary Fig. 8).

Methods were repeated using 40-year time periods (e.g., 2061–2100 vs. 1981–2020), which may be more representative of observational records available to practitioners in some regions and less susceptible to non-stationarity. At the global scale, magnitudes of projected change resemble those derived from 80-year periods but with lower SNRs resulting in fewer grids exhibiting unprecedented change (Supplementary Table 6; Supplementary Fig. 9). Reduced detection of signals among noise may be influenced, in part, by criteria set in our definition of categorical SNRs where the climate change projection must exceed all changes calculated using piControl simulations (see Methods). Such criteria are subjective and the use of alternate criteria[34] may have a different effect on the consistency between 40- and 80-year periods, however, this was not tested. Nonetheless, we see greater magnitudes and more unprecedented change for the latter segment of the century (2061–2100) compared to an earlier segment (2021–2060). We also postulate that the use of longer records, when available, is favored among many practitioners for drought planning in high-risk circumstances.

Methods were also repeated to explore projected changes in the most severe and longest-duration droughts (denoted as SEVx and DURx; see Supplementary Discussion). Due to the rare nature of these events, results must be interpreted with extra caution but may be of special interest to water managers and planners. Projections (Supplementary Fig. 7 and Supplementary Table 5) illustrate similar spatial patterns to mean drought (SEV and DUR) but correspond to smaller areas with unprecedented change. In cases where projections are indistinguishable from internal climate variability, one should keep in mind that it is not uncommon to find droughts twice as severe or five years longer than those in the 1941–2020 period simply due to internal climate variability. In these cases, a future projection that is indistinguishable from internal variability may still be of special relevance. The ensemble interquartile range of future projections (Supplementary Table 4; Supplementary Table 5) helps reveal a portion of uncertainty.

DISCUSSION

Multi-century simulations and improved land surface representations in CMIP5 GCMs create opportunities to explore changes in future multi-year hydrologic drought and to compare those changes against internal climate variability. Using 640-year scenarios of simulated runoff derived from an ensemble of eighteen GCMs, two future emissions scenarios, pseudo-observed runoff products, and a common bias adjustment technique, we provide a global reference that can support adaptation discussions, inclusive of regions that may otherwise be data sparse. Furthermore, our analysis provides a basis for future work in understanding climate change projections of multi-year hydrologic drought.

Results show that GCM skill in reproducing observed runoff conditions varies by model and region but is less variable among multiple simulations from the same institution. Furthermore, GCMs can perform comparably to an alternate pseudo-observed runoff product suggesting that work remains in resolving observational uncertainties. Quantile mapping (QM) effectively reduces bias in the mean and distribution of GCM simulated runoff, but biases in autocorrelation and the average number of cumulative dry years remain. Repeating our approach without QM

reveals similar spatial patterns but with slightly stronger signals (Supplementary Fig. 6), suggesting that the bias adjustment process did not wildly influence findings.

Results identify wetting and drying regions that are consistent with previous studies of sub-annual to annual runoff but provide a multi-year drought perspective which can be important for water managers and long-term planning. Spatial patterns in changing MAR generally align with spatial patterns in changing SEV and DUR, but areas with unprecedented changes in mean runoff do not always align with unprecedented changes in drought. For example, over portions of India, unprecedented increases in MAR correspond to decreases in SEV and DUR that are classified as indistinguishable (from internal climate variability) due to lower SNRs.

For regions with unreliable or indistinguishable projections, our work provides a basis to inform limits for scenario planning and sensitivity analyses and identifies areas where a fulsome understanding of internal variability (e.g., through paleoclimate studies) may be of value. In regions with unprecedented projections, our findings may contribute evidence in support of climate adaptation decisions, such as adjusting reservoir operating rules or increasing turbine efficiency[52–54]. We also find fewer and lower magnitude, unprecedented hydrologic changes by avoiding a future with greater greenhouse gas emissions.

Following a similar framework, additional datasets (hydrologic model ensembles, reference products), along with more advanced bias adjustment techniques and different indices (skill and drought), can provide additional evidence to support adaptation considerations. Future studies may also examine outputs from more recent climate models (CMIP6[55]) that incorporate improved hydrologic representations in their LSMs[19], however, results may not wildly deviate from CMIP5 findings[15,43]. A further look into GCMs during severe drought periods may shed light on what drives multi-year hydrologic drought to persist in time (e.g., oceanic teleconnections, radiative forcing, or feedback loops). This study omits stochastic approaches for drought risk analysis, however, the 640-year scenarios generated could complement stochastic methods[56] and help researchers understand how non-stationarity impacts drought risk estimation. Utilization of longer time series from GCMs (past1000 experiment[18] or large ensembles[57–59]) may help reduce sampling uncertainty.

Our study neglects the impacts of routing and anthropogenic effects (e.g., reservoir regulation, diversions, land use change, and water demands), which can ease or exacerbate drought impacts, are embedded in observational runoff datasets, and contribute to challenges associated with non-stationarity[60]. Some anthropogenic effects, such as reservoir regulation, are typically pertinent at sub-annual time scales[61] but can play an important role in managing multi-year drought risk. Other anthropogenic effects, such as human water consumption for growing populations and agricultural withdrawals, only intensify societal impacts, especially in more vulnerable developing countries[62]. Incorporation of anthropogenic effects, such as reservoir regulation, would require runoff inputs with finer temporal resolution (i.e., sub-annual) and the development of demand, routing, and regulation models either offline or directly coupled in future generations of climate models[63]. Although such considerations were not incorporated herein, we recognize that anthropogenic factors can have a compounding influence on local drought impact assessments and are a topic for further study. Uncertainty remains (even for regions with unprecedented change) and further work is required[21] prior to the establishment of climate model runoff as a high-confidence, design-suitable variable. However, we recognize that lower confidence variables can still offer pragmatic value for adaptation planning and risk analysis[64].

METHODS

Pseudo-observed gridded runoff data

Monthly gridded runoff derived using the Linear Optimal Runoff Aggregate method (LORA[36]) was selected as the observed reference data for GCM evaluation and bias adjustment. Data were available from 1980 to 2012 globally at a 0.5° grid and represent a merging of eleven gridded runoff products[65], weighted by performance. LORA performed better than ten individual hydrologic model products[66] and the variable infiltration capacity (VIC) global gridded runoff data[67]. Monthly LORA data were aggregated to annual runoff by calendar year and upscaled (by averaging four 0.5° grids) to a 1° grid for closer comparison with GCM data. No routing was performed, given the focus of our study was on MAR.

Runoff from GRUN[37] and the Global Runoff Data Center (GRDC)[68] datasets were also used in a complementary fashion. GRUN achieves an extended time period (1902–2014) by coupling meteorological forcing with machine learning and was upscaled to a 1° grid to serve as a second gridded runoff dataset for evaluation of model skill. GRDC provides streamflow observations and spatially averaged runoff climatologies derived from the combination of streamflow records and a water balance model. GRDC streamflow observations are used to aid the visual comparison of bias-adjusted runoff time series. However, since time series of GRDC gridded runoff are not available, these data were not applicable for skill evaluation or bias adjustment. Individual models, on which LORA is based, provide monthly estimation of the global terrestrial water budget but may utilize shorter time periods (e.g., 1984–2010[67]) and were outperformed by LORA[36]. Runoff data from reanalysis products were considered but found to propagate precipitation errors into hydrologic errors, have imperfect physical representation[69], and have poor performance when compared to terrestrial biosphere models[70]. The National Oceanic and Atmospheric Administration's 20th-century reanalysis with 1° global resolution is attractive for its long duration (1836–2015)[71], but its runoff output has yet to be fully evaluated.

GCM runoff data

Total runoff (surface and sub-surface; mrro) data were obtained from GCMs contributing to the Coupled Model Intercomparison Project Phase 5 (CMIP5). GCMs were included if they contained monthly runoff data for a historic simulation (1861–2005), future simulations (2006–2100) driven by two atmospheric forcing scenarios (RCP4.5 and RCP8.5)[72] and a pre-industrial control simulation (piControl)[18] containing at least 500 years of data under idealized stationary atmospheric composition. The availability of multi-century time series plays an important role in the sampling of droughts and identification of climate signals, which is discussed later.

Eighteen GCMs met these criteria (Supplementary Table 1); seven of which had archived runoff data for multiple runs in the historic and future periods representing different GCM initial conditions. Multi-run ensembles were included to sample GCM internal variability, but a one-model-one-vote approach was used for ensemble statistics to help maintain a democratic ensemble[73] and to avoid skewing ensemble statistics toward GCMs with multiple runs.

GCMs are comprised of components that simulate different aspects of the Earth's environment. Typical components include the atmosphere, the ocean, an LSM, sea-ice, as well as atmospheric chemistry and carbon cycling for Earth System Models[74]. The LSM represents land cover, soil, and vegetation and is important for computing the flux of energy and water between the atmosphere and land. LSMs are similar to hydrologic models but follow a more physically-based approach by solving both water and energy balances[25,66]. It's important that the LSM is

dynamically coupled as latent heat fluxes can affect atmospheric conditions, and freshwater can affect oceanic conditions. Some GCMs have special names for their LSM, while others simply lump the land surface scheme with the atmospheric model (Supplementary Table 1).

LSMs vary in complexity[19] and can influence future projections[75]. Generally, each grid is characterized by one or more land cover types which may include bare soil and vegetation[76], snow-covered[77], and glacier, wetland, lake, and urban[78]. Vegetation may be further categorized by plant function type (PFT), including coniferous trees, deciduous trees, grasses, crops, shrubs, and rainforests. PFTs are important hydrologically (e.g., they influence water retention in the vegetative canopy, temperature, and evapotranspiration) and biologically as vegetation interacts with atmospheric carbon. Snow is typically modeled by one of four schemes ranging in complexity from a simple approach that uniformly distributes snow within a grid to a more sophisticated scheme that resolves sub-grid snow cover and allows multiple snowpack layers[79]. Beyond vegetation and snow, LSMs vary in how they handle lakes, lake ice, soil properties, drainage, routing, and other processes. It is easy to saturate in individual LSM detail (see literature cited in Supplementary Table 1), but a sample of CMIP5 GCMs[80] shows that LSM configuration can range from 3 to 20 soil layers, representing 3–47.3 m depths and utilize 5–15 PFTs.

Since GCMs operate with individual spatial resolutions and land masks, a common grid and land mask were derived to provide a consistent basis for evaluation and inter-comparison. GCM land masks and runoff data were re-gridded to correspond with the upscaled LORA 1° grid using the nearest neighbor approach, which helps preserve the GCM's original 'coarseness' of land features and total runoff volume. For large basins, different interpolation techniques have little effect on analyses[35]. Land masks can be binary or mosaic. In our study, mosaic land masks are replaced by 100 when the grid is ≥50% land and 0 otherwise. The final mask is produced by selecting common land grids among all 18 GCMs and upscaled LORA data. Excluding Antarctica (not included in LORA), the final mask contained 11,939 land grids (approximately 108,817,028 km^2), or ~85% of the total land area available in LORA. The result is a coarse world map (Fig. 1) resulting from different land mask-resolution combinations used across the GCM ensemble. Several geographic features are notably distorted (e.g., Great Lakes) or are completely removed (e.g., the Italian Peninsula).

Skill evaluation

Skill evaluation focused only on GCM mean annual runoff and was evaluated against LORA and GRUN data across a 33-year period (1980–2012). Papers cited in our results and Supplementary References provide additional context on other related variables.

Six indices were selected to evaluate GCM skill. Four common volume-based indices include MAR (mm), minimum of mean annual runoff (MIN; mm), maximum of mean annual runoff (MAX; mm), and coefficient of variation of mean annual runoff (CV; unitless; standard deviation divided by MAR). One-year lagged autocorrelation (AC1; unitless) and the average number of cumulative dry years (CDY; years) are included as two additional indices to provide context beyond typical statistical properties[81]. Collectively, these six indices provide information relevant to drought definition (MAR), simulation of individual extreme dry/wet years (MIN and MAX), the variability in year-to-year runoff (CV), how well persistence is simulated (AC1), and whether a model is capable of simulating extended periods of dry conditions (CDY). Comparison of time series correlation (model vs. observed) was not meaningful since GCMs are not designed to model observed sequences of events.

Established statistical metrics[38] were adopted herein to provide an overall indication of GCM skill in space, however, we do not integrate skill monthly. Instead, a single (annual) value was computed for each index to evaluate GCM skill on an annual climatology as opposed to sub-annual cycles[39]. Root mean square error (RMSE) was used to gauge skill and was supplemented with relative RMSE ($RMSE_R$) to provide a closer look at a model's performance with respect to the median performing model in the ensemble. $RMSE_{R\text{-}All}$ (mean of $RMSE_R$ for all six indices) was computed as an overall score for each GCM simulation: The best (worst) performing simulation, overall, will have the lowest (greatest) $RMSE_{R\text{-}All}$.

While skill was assessed for all GCM simulations, results were shown for one run per model (plus additional perturbed physics runs for GISS models). Because skill is evaluated for a period spanning two GCM experiments (historic up to 2005 and RCP8.5 after 2005), skill evaluation merged 26 years from the GCM historic simulation and 7 years from the RCP8.5 simulation. Using one RCP provided a more concise summary and, in most cases, did not dramatically impact model skill. Special consideration was made for two cases to avoid unrealistic representation of skill: (1) GCM grids with a MAR of zero were assigned a CV of zero to avoid division by zero; and (2) GCM grids with equal runoff over the entire 1980–2012 period (occurs in some GCMs over dry areas, like northern Africa), AC1 and CDY were assigned a value of zero.

Bias adjustment

A common univariate QM approach was used for bias adjustment that was analogous to the daily translation[82], QMv1a[83], and 1Dqm[84] methods. A single QM (transfer function) was developed for each GCM grid in a simulation by comparing GCM simulated runoff to the LORA reference data. QMs, derived from two non-parametric cumulative distribution functions (CDFs), were generated and then applied to correct the entire GCM simulation. Although bias adjustment has been most often performed on temperature and precipitation, it has been applied to other climate model variables[42,85–87] and has been used in streamflow forecasting to improve forecast skill[88]. The application of QM to climate model runoff[89,90] is not widely published but is a practical extension of existing methods to adjust the volume bias of annual runoff estimation.

The application of a year-round QM can be suitable for annual data[91] and aligns with our objective of exploring multi-year hydrologic drought. While finer temporal scales and a larger number of transfer functions (e.g., one per month or one per day of the year) are common for the bias adjustment of meteorological variables[82,84], this adds unnecessary complexity to our study.

QMs were derived from 33 data points (i.e., one for each year, 1980–2012), and adjustment values were assessed at 33 fixed, equally spaced percentiles. Adjustments were linearly interpolated between fixed percentiles and held constant for GCM runoff values beyond the minimum and maximum GCM values[84] in the 1980–2012 period. The computational process first set all occurrences of negative runoff to zero, then applied additive adjustment, and reset residual negative runoff to zero once again. This process aligned with physical consistency in the LORA dataset (i.e., runoff bound at zero) and avoided exaggerated multiplicative adjustments that emerge when near-zero values are present in a GCM's reference time series and absent from the LORA dataset.

The entire simulation (piControl, historic, and future data) is bias-adjusted using the same QM. Steps to preserve the original climate model trend are performed in some studies, but there is no consensus regarding this step[92,93]. Others[84,91] removed, then reapplied temperature regression after QMing to preserve the climate model's long-term signal but did not use this procedure to force preservation of precipitation trends. Since runoff trends may be quantile dependent, and a uniform approach to account for trends (e.g., linear or polynomial) may not be applicable for every

grid in our domain, no trend preservation procedure was applied herein. This remains a topic for future exploration. Compared to alternate bias adjustment approaches such as the delta method[83], our selected approach improves the preservation of the GCM's original sequencing of events, which is an important feature for assessing changes to multi-year drought events.

Special attention was paid to LORA and GCM grids with ≥75% repetitive runoff values during the 1980–2012 period. This condition, found in dry regions, can cause issues with the rank-based ordering in the QM scheme, which applies small, random perturbations (on the order of 1^{-12} mm/year) to avoid singularities in CDFs. These grids were flagged and bias-adjusted with a single mean adjustment rather than a quantile-dependent adjustment. At any grid, no more than eight GCMs exhibited such conditions. Without special attention, these conditions could unintentionally allow random perturbations (purely stochastic) to influence GCM-simulated droughts. Steps to address special cases may somewhat limit analysis in affected regions as only the mean state is bias-adjusted rather than the entire distribution. Special considerations are common to account for distinct circumstances. For example, parameter checks have been used to fix unrealistic (e.g., negative) values[42].

Assessing changes in multi-year hydrologic drought

We use bias-adjusted MAR time series to compare droughts in different time periods (i.e., piControl, historic, and future; Supplementary Table 7). PiControl years are arbitrarily assigned nominal dates such that the end of the piControl period aligns with the beginning of the historic period, but in accordance with methods to quantify model agreement, the first 100 years are omitted[34]. PiControl (1461–1860), historic (1861–2005), and future (2006–2100) experiments are assessed in time slices of 80 years. Projected changes (deltas; Δ) compare values in a future time slice (2021–2100) to a reference time slice (1941–2020). The remaining years (1461–1940) are used to quantify internal climate variability and designated "PR" (precede the reference time slice). For piControl years (1461–1860), a moving window is used to sample 321 80-year time slices. This moving window captures drought sequences that may have otherwise been missed when using discrete time slices. Because of potential discontinuity between CMIP5's piControl and historic experiments, 1861–1940 is included as a discrete-time slice for a total of 322 PR time slices.

Four hydrologic characteristics were explored. Changes in MAR and CV are indicative of average annual water supply and year-to-year runoff variability. Change in SEV represents the accumulated volume of water deficit (difference between threshold runoff and simulated runoff) experienced in a typical drought, summed for all years of each drought. SEV is expressed as a positive value with larger numbers representing greater water deficits. Finally, change in DUR provides information regarding the temporal extent to which drought conditions may persist for a typical event. DUR is expressed as the count of years where simulated runoff is below the threshold runoff. Droughts of similar severity can coincide with different duration, which can result in distinct socio-economic circumstances. The shortest possible drought duration was one year, in which case, duration (1) and severity are assigned to that year. For multi-year events, duration and severity are assigned to the first year of the drought, and if a drought spans multiple discrete time slices, one drought event is recorded per time slice instead of one large event.

Following the theory of runs[94], a drought begins and then ends when MAR crosses a threshold defined as the GCM's MAR during 1941–2020. Our time slices exceed the typical period (≤30 years) used in climate change studies but are selected to better suit the assessment of multi-year events[95] and reduce sensitivity to low-frequency climate variability[1], which could result in multi-decadal periods of relatively wet or dry conditions. In regions subject to non-stationary hydrologic conditions, we caution that future projected changes may not be representative of conditions near 2021 and 2100. To help interpretation of our findings, methods were repeated using 40-year time slices (Supplementary Discussion, Supplementary Fig. 9, and Supplementary Table 6). For global analysis, droughts were analyzed grid-by-grid. For river basin analyses, annual runoff in all grids within the basin was averaged and hydrologic characteristics were computed on the basin-aggregated runoff time series.

Ensemble median projections, categorical SNR, and agreement on the direction of change were used to summarize the results. These three pieces of information provide evidence that separates the lack of significant signal from the lack of agreement on the sign of the signal[96]. For GCMs with multiple runs and physics experiments, the intra-model median projection is first determined, and then ensemble statistics (e.g., median) among all GCMs are computed. This approach avoided skewing ensemble statistics towards GCMs with multiple runs. For the assessment of future projections, RCP4.5 and RCP8.5 were evaluated together as equiprobable scenarios (i.e., one ensemble consisting of 18 GCMs and 2 RCPs for a total of 36 simulations) as well as independently.

Categorical SNRs identify future projections (signals) that fall beyond simulated internal climate variability (noise). For each GCM simulation, grid/basin, and hydrologic characteristic, deltas (Δ) were computed for the future time slice and for the 322 PR time slices relative to the reference time slice (e.g., $\Delta MAR_{PR1} = MAR_{PR1} - MAR_{Reference}$). From the 322 PR deltas, the maximum and minimum were computed (ΔMAR_{PR_MAX} and ΔMAR_{PR_MIN}) for comparison against ΔMAR_{Future}. This process is similar to other approaches[97–100]. For each GCM simulation, SNR is assigned a value of 1 when ΔMAR_{Future} is beyond ΔMAR_{PR_RANGE} [ΔMAR_{PR_MIN}; ΔMAR_{PR_MAX}] and 0 otherwise. For GCMs with multiple runs and physics experiments, an intra-model median SNR is computed.

Ensemble agreement and SNRs were interpreted together: Findings are characterized as unprecedented when >50% of the projections have SNR of 1 and 80% or more of those projections agree on the direction of change (shown with color and hatching on maps). An unreliable finding is associated with >50% of projections having an SNR of 1 but <80% of those agreeing on the direction of change (shown in white on maps). Projections are characterized as indistinguishable from internal climate variability when 50% or more of projections have an SNR of 0 (shown in color on maps). While similar interpretations have been described using the term robust[34,96], we've selected alternate terms to cautiously avoid an overprescription of confidence in projected changes.

DATA AVAILABILITY

Global Climate Model runoff data used during this study are openly available from the Earth System Grid Federation (ESGF) at https://esgf.llnl.gov/. Other data sources are cited in-text. Additional data generated during this study are presented in the Supplementary Information or available upon reasonable request to the authors.

CODE AVAILABILITY

The code used to process and analyze data is described in Methods and available upon reasonable request to the authors.

Received: 18 March 2023; Accepted: 11 October 2023;
Published online: 07 November 2023

REFERENCES

1. Trenberth, K. E. et al. Global warming and changes in drought. *Nat. Clim. Change* **4**, 17–22 (2014).
2. Bifulco, M. & Ranieri, R. Impact of drought on human health. *Eur. J. Intern. Med.* **46**, e9–e10 (2017).
3. Yusa, A. et al. Climate Change, drought and human health in Canada. *Int. J. Environ. Res. Public Health* **12**, 8359–8412 (2015).

4. United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development. A/RES/70/1 (2015).
5. O'Connell, M., Voisin, N., Macknick, J. & Fu, T. Sensitivity of Western U.S. power system dynamics to droughts compounded with fuel price variability. *Appl. Energy* **247**, 745–754 (2019).
6. van Dijk, A. I. J. M. et al. The Millennium Drought in southeast Australia (2001–2009): natural and human causes and implications for water resources, ecosystems, economy, and society. *J. Water Resour. Res.* **49**, 1040–1057 (2013).
7. Smith, A. B. & Matthews, J. L. Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. *Nat. Hazards* **77**, 1829–1851 (2015).
8. Hoerling, M. et al. Anatomy of an extreme event. *J. Clim.* **26**, 2811–2832 (2013).
9. Carrão, H., Naumann, G. & Barbosa, P. Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability. *Glob. Environ. Change* **39**, 108–124 (2016).
10. Alkama, R., Marchand, L., Ribes, A. & Decharme, B. Detection of global runoff changes: results from observations and CMIP5 experiments. *Hydrol. Earth Syst. Sci.* **17**, 2967–2979 (2013).
11. Arnell, N. W. & Gosling, S. N. The impacts of climate change on river flow regimes at the global scale. *J. Hydrol.* **486**, 351–364 (2013).
12. Touma, D., Ashfaq, M., Nayak, M. A., Kao, S.-C. & Diffenbaugh, N. S. A multi-model and multi-index evaluation of drought characteristics in the 21st century. *J. Hydrol.* **526**, 196–207 (2015).
13. Ukkola, A. M. et al. Evaluating CMIP5 model agreement for multiple drought metrics. *J. Hydrometeorol.* **19**, 969–988 (2018).
14. Spinoni, J. et al. Future global meteorological drought hot spots: a study based on CORDEX data. *J. Clim.* **33**, 3635–3661 (2020).
15. Cook, B. I. et al. Twenty-first century drought projections in the CMIP6 forcing scenarios. *Earth's Fut.* **8**, e2019EF001461 (2020).
16. Knutti, R. & Sedláček, J. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change* **3**, 369–373 (2013).
17. Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J. & Bierkens, M. F. P. On the suitability of GCM runoff fields for river discharge modeling: a case study using model output from HadGEM2 and ECHAM5. *J. Hydrometeorol.* **13**, 140–154 (2012).
18. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
19. Clark, M. P. et al. Improving the representation of hydrologic processes in Earth System Models. *J. Water Resour. Res.* **51**, 5929–5956 (2015).
20. Davison, B. et al. What is missing from the prescription of hydrology for land surface schemes? *J. Hydrometeorol.* **17**, 2013–2039 (2016).
21. Lehner, F. et al. The potential to reduce uncertainty in regional runoff projections from climate models. *Nat. Clim. Change* **9**, 926–933 (2019).
22. Wood, A. W., Lettenmaier, D. P. & Palmer, R. N. Assessing Climate Change Implications for Water Resources Planning. *Clim. Change* **37**, 203–228 (1997).
23. Chen, J., Brissette, F. P., Poulin, A. & Leconte, R. Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *J. Water Resour. Res.* **47**, W12509 (2011).
24. Hagemann, S. et al. Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dyn.* **4**, 129–144 (2013).
25. Haddeland, I. et al. Multimodel estimate of the global terrestrial water balance: setup and first results. *J. Hydrometeorol.* **12**, 869–884 (2011).
26. Haddeland, I. et al. Global water resources affected by human interventions and climate change. *Proc. Natl Acad. Sci. USA* **111**, 3251 (2014).
27. Giuntoli, I., Vidal, J. P., Prudhomme, C. & Hannah, D. M. Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models. *Earth Syst. Dyn.* **6**, 267–285 (2015).
28. Prudhomme, C. et al. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Natl Acad. Sci. USA* **111**, 3262–3267 (2014).
29. van Huijgevoort, M. H. J., van Lanen, H. A. J., Teuling, A. J. & Uijlenhoet, R. Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge. *J. Hydrol.* **512**, 421–434 (2014).
30. Arora, V. K., Chiew, F. H. S. & Grayson, R. B. A river flow routing scheme for general circulation models. *J. Geophys. Res. Atmos.* **104**, 14347–14357 (1999).
31. Reick, C. H., Raddatz, T., Brovkin, V. & Gayler, V. Representation of natural and anthropogenic land cover change in MPI-ESM. *J. Adv. Model. Earth Syst.* **5**, 459–482 (2013).
32. Anderegg, W. R. L., Trugman, A. T., Bowling, D. R., Salvucci, G. & Tuttle, S. E. Plant functional traits and climate influence drought intensification and land–atmosphere feedbacks. *Proc. Natl Acad. Sci. USA* **116**, 14071 (2019).
33. Milly, P. C. D., Dunne, K. A. & Vecchia, A. V. Global pattern of trends in streamflow and water availability in a changing climate. *Nature* **438**, 347–350 (2005).
34. Collins, M. et al. Long-term climate change: projections, commitments and irreversibility. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, *1029–1136* (Cambridge University Press, 2013).
35. Koirala, S., Hirabayashi, Y., Mahendran, R. & Kanae, S. Global assessment of agreement among streamflow projections using CMIP5 model outputs. *Environ. Res. Lett.* **9**, 11 (2014).
36. Hobeichi, S., Abramowitz, G., Evans, J. & Beck, H. E. Linear Optimal Runoff Aggregate (LORA): a global gridded synthesis runoff product. *Hydrol. Earth Syst. Sci.* **23**, 851–870 (2019).
37. Ghiggi, G., Humphrey, V., Seneviratne, S. I. & Gudmundsson, L. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data* **11**, 1655–1674 (2019).
38. Gleckler, P. J., Taylor, K. E. & Doutriaux, C. Performance metrics for climate models. *J. Geophys. Res. Atmos* **113**, D06104 (2008).
39. Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **118**, 1716–1733 (2013).
40. Eyring, V. et al. ESMValTool (v1.0)—a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev.* **9**, 1747–1802 (2016).
41. Li, L. et al. Evaluating Global Land Surface Models in CMIP5: analysis of ecosystem water- and light-use efficiencies and rainfall partitioning. *J. Clim.* **31**, 2995–3008 (2018).
42. Papadimitriou, L. V., Koutroulis, A. G., Grillakis, M. G. & Tsanis, I. K. The effect of GCM biases on global runoff simulations of a land surface model. *Hydrol. Earth Syst. Sci.* **21**, 4379–4401 (2017).
43. Guo, H., Zhan, C., Ning, L., Li, Z. & Hu, S. Evaluation and comparison of CMIP6 and CMIP5 model performance in simulating the runoff. *Theor. Appl. Climatol.* **149**, 1451–1470 (2022).
44. Rodell, M. et al. Emerging trends in global freshwater availability. *Nature* **557**, 651–659 (2018).
45. Feng, H. & Zhang, M. Global land moisture trends: drier in dry and wetter in wet over land. *Sci. Rep.* **5**, 18018 (2015).
46. Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res. Atmos.* **118**, 2473–2493 (2013).
47. Mankin, J. S., Seager, R., Smerdon, J. E., Cook, B. I. & Williams, A. P. Mid-latitude freshwater availability reduced by projected vegetation responses to climate change. *Nat. Geosci.* **12**, 983–988 (2019).
48. Greve, P. et al. Global assessment of trends in wetting and drying over land. *Nat. Geosci.* **7**, 716–721 (2014).
49. Lehner, F. et al. Projected drought risk in 1.5°C and 2°C warmer climates. *Geophys. Res. Lett.* **44**, 7419–7428 (2017).
50. Cook, B. I., Anchukaitis, K. J., Touchan, R., Meko, D. M. & Cook, E. R. Spatiotemporal drought variability in the Mediterranean over the last 900 years. *J. Geophys. Res. Atmos.* **121**, 2060–2074 (2016).
51. Sperna Weiland, F. C., van Beek, L. P. H., Kwadijk, J. C. J. & Bierkens, M. F. P. Global patterns of change in discharge regimes for 2100. *Hydrol. Earth Syst. Sci.* **16**, 1047–1062 (2012).
52. Arsenault, R., Brissette, F., Malo, J.-S., Minville, M. & Leconte, R. Structural and non-structural climate change adaptation strategies for the Péribonka water resource system. *Water Resour. Manag.* **27**, 2075–2087 (2013).
53. van Vliet, M. T. H., Wiberg, D., Leduc, S. & Riahi, K. Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nat. Clim. Change* **6**, 375–380 (2016).
54. Turner, S. W. D., Ng, J. Y. & Galelli, S. Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model. *Sci. Total Environ.* **590-591**, 663–675 (2017).
55. Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
56. Slater, L. J. et al. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrol. Earth Syst. Sci.* **25**, 3897–3935 (2021).
57. Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change* **2**, 775–779 (2012).
58. Kay, J. E. et al. The Community Earth System Model (CESM) Large Ensemble Project: a community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteorol. Soc.* **96**, 1333–1349 (2015).
59. Oudar, T., Kushner, P. J., Fyfe, J. C. & Sigmond, M. No impact of anthropogenic aerosols on early 21st century global temperature trends in a large initial-condition ensemble. *Geophys. Res. Lett.* **45**, 9245–9252 (2018).

60. Van Loon, A. F. et al. Streamflow droughts aggravated by human activities despite management. *Environ. Res. Lett.* **17**, 044059 (2022).
61. Déry, S. J., Stadnyk, T. A., MacDonald, M. K., Koenig, K. A. & Guay, C. Flow alteration impacts on Hudson Bay river discharge. *Hydrol. Process.* **32**, 3576–3587 (2018).
62. AghaKouchak, A. et al. Anthropogenic drought: definition, challenges, and opportunities. *Rev. Geophys.* **59**, e2019RG000683 (2021).
63. Nazemi, A. & Wheater, H. S. On inclusion of water resource management in Earth system models—Part 2: representation of water supply and allocation and opportunities for improved modeling. *Hydrol. Earth Syst. Sci.* **19**, 63–90 (2015).
64. Cannon, A. J., Jeong, D. I., Zhang, X. & Zwiers, F. W. Climate-resilient buildings and core public infrastructure—an assessment of the impact of climate change on climatic design data in Canada. *Env. Clim. Change Canada* En4-415/2020E-PDF (2020).
65. Schellekens, J. et al. A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset. *Earth Syst. Sci. Data* **9**, 389–413 (2017).
66. Beck, H. E. et al. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.* **21**, 2881–2903 (2017).
67. Zhang, Y. et al. A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. *Hydrol. Earth Syst. Sci.* **22**, 241–263 (2018).
68. Fekete, B. M., Vörösmarty, C. J. & Grabs, W. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Glob. Biogeochem. Cycles* **16**, 15-11-15-10 (2002).
69. Reichle, R. H. et al. Assessment and enhancement of MERRA land surface hydrology estimates. *J. Clim.* **24**, 6322–6338 (2011).
70. Schwalm, C. R. et al. A model–data intercomparison of simulated runoff in the contiguous United States: results from the North America Carbon Regional and Continental Interim-Synthesis. *Biogeosci. Discuss.* **2014**, 1801–1826 (2014).
71. Slivinski, L. C. et al. Towards a more reliable historical reanalysis: improvements for version 3 of the twentieth century reanalysis system. *Q. J. R. Meteorol. Soc.* **145**, 2876–2908 (2019).
72. van Vuuren, D. P. et al. The representative concentration pathways: an overview. *Clim. Change* **109**, 5 (2011).
73. Sanderson, B. M., Knutti, R. & Caldwell, P. A representative democracy to reduce interdependency in a multimodel ensemble. *J. Clim.* **28**, 5171–5194 (2015).
74. Flato, G. et al. Evaluation of climate models. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, *741–866* (Cambridge University Press, 2013).
75. Yu, M., Wang, G. & Chen, H. Quantifying the impacts of land surface schemes and dynamic vegetation on the model dependency of projected changes in surface energy and water budgets. *J. Adv. Model. Earth Syst.* **8**, 370–386 (2016).
76. Rosenweig, C. & Abramopoulos, F. Land-surface model development for the GISS GCM. *J. Clim.* **10**, 2040–2054 (1997).
77. Verseghy, D. L. The Canadian land surface scheme (CLASS): its history and future. *Atmosphere* **38**, 1–13 (2000).
78. Oleson, K. W. et al. Technical Description of version 4.0 of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-478+STR. National Center for Atmospheric Research (2010).
79. Thackeray, C. W., Fletcher, C. G. & Derksen, C. Quantifying the skill of CMIP5 models in simulating seasonal albedo and snow cover evolution. *J. Geophys. Res. Atmos.* **120**, 5831–5849 (2015).
80. Huang, Y., Gerber, S., Huang, T. & Lichstein, J. W. Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data. *Glob. Biogeochem. Cycles* **30**, 1827–1846 (2016).
81. Papalexiou, S. M., Rajulapati, C. R., Clark, M. P. & Lehner, F. Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. *Earth's Fut.* **8**, e2020EF001667 (2020).
82. Mpelasoka, F. S. & Chiew, F. H. S. Influence of rainfall scenario construction methods on runoff projections. *J. Hydrometeorol.* **10**, 1168–1183 (2009).
83. Themeßl, M. J., Gobiet, A. & Heinrich, G. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Clim. Change* **112**, 449–468 (2012).
84. Gennaretti, F., Sangelantoni, L. & Grenier, P. Toward daily climate scenarios for Canadian Arctic coastal zones with more realistic temperature-precipitation interdependence. *J. Geophys. Res. Atmos.* **120**, 11,862–811,877 (2015).
85. Wilcke, R. A. I., Mendlik, T. & Gobiet, A. Multi-variable error correction of regional climate models. *Clim. Change* **120**, 871–887 (2013).
86. Grenier, P. Two types of physical inconsistency to avoid with univariate quantile mapping: a case study over North America concerning relative humidity and its parent variables. *J. Appl. Meteorol. Climatol.* **57**, 347–364 (2018).
87. Cheng, C. S., Lopes, E., Fu, C. & Huang, Z. Possible impacts of climate change on wind gusts under downscaled future climate conditions: updated for Canada. *J. Clim.* **27**, 1255–1270 (2014).
88. Hashino, T., Bradley, A. A. & Schwartz, S. S. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol. Earth Syst. Sci.* **11**, 939–950 (2007).
89. Duong, D. T., Tachikawa, Y. & Yorozu, K. Application of a land surface model for bias correction of runoff generation data from MRI-AGCM3.2S Dataset. In: *THA 2015 International Conference on Climate Change and Water & Environment Management in Monsoon Asia (Bangkok, Thailand, 2015)*.
90. Manee, D., Tachikawa, Y., Ichikawa, Y. & Yorozu, K. Evaluation of bias correction methods for future river discharge projection. *J. Jpn. Soc. Civ.* **72**, I_7–I_12 (2016).
91. Grenier, P., Firlej, A., Blondlot, A., Logan, T. & Ricard, M.-P. The issue of properly ordering climate indices calculation and bias correction before identifying spatial analogs for agricultural applications. *Clim. Serv.* **16**, 100122 (2019).
92. Hempel, S., Frieler, K., Warszawski, L., Schewe, J. & Piontek, F. A trend-preserving bias correction—the ISI-MIP approach. *Earth Syst. Dyn.* **4**, 219–236 (2013).
93. Gobiet, A., Suklitsch, M. & Heinrich, G. The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal. *Hydrol. Earth Syst. Sci.* **19**, 4055–4066 (2015).
94. Yevjevich V. An objective approach to definitions and investigations on continental hydrologic droughts In: *Hydrology Papers*. (Colorado State University, 1967).
95. Svoboda, M., Hayes, M. & Wood, D. Standardized Precipitation Index User Guide. World Meteorological Organization WMO-No.1090 (2012).
96. Tebaldi, C., Arblaster, J. M. & Knutti, R. Mapping model agreement on future climate projections. *Geophys. Res. Lett.* **38**, L23701 (2011).
97. New, M. & Hulme, M. Representing uncertainty in climate change scenarios: a Monte-Carlo approach. *Integr. Assess.* **1**, 203–213 (2000).
98. Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. African climate change 1900-2100. *Clim. Res.* **17**, 145–168 (2001).
99. Santer, B. D. Separating signal and noise in atmospheric temperature changes: the importance of timescale. *J. Geophys. Res. Atmos.* **116**, D22105 (2011).
100. Perkins-Kirkpatrick, S. E., Fischer, E. M., Angélil, O. & Gibson, P. B. The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.* **12**, 044005 (2017).
101. Lehner, B. et al. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front. Ecol. Evol.* **9**, 494–502 (2011).

ACKNOWLEDGEMENTS

The authors acknowledge the World Climate Research Program's Working Group on Coupled Modeling, which is responsible for CMIP, and thank the climate modeling groups (listed in Supplementary Table 1) for producing and making available their model output. The authors also acknowledge Dr. Patrick Grenier of the Ouranos Consortium for bias adjustment methodology guidance and review.

AUTHOR CONTRIBUTIONS

Experiment design, analysis, interpretation of results, and paper writing by M.J.F.V. Results interpretation, methodological guidance, review, and paper edits by T.A.S.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00496-y.

Correspondence and requests for materials should be addressed to Michael J. F. Vieira.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.