Incorporating Multiyear Temperature Predictions for Water Resources Planning

ERIN TOWLER^{a,b} AND DAVID YATES^a

^a *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 10 June 2020, in final form 10 November 2020)

ABSTRACT: Multiyear climate predictions provide climate outlooks from years to a decade in advance. As multiyear temperature predictions become more mainstream and skillful, guidance is needed to assist practitioners who wish to explore this maturing field. This paper demonstrates the process and considerations of incorporating multiyear temperature predictions into water resources planning. Multiyear temperature predictions from the Community Earth System Model Decadal Prediction Large Ensemble are presented as discrete and probabilistic products and are used to force two common hydrologic modeling approaches: conceptual and empirical. The approaches are demonstrated to simulate streamflow in the upper Colorado River basin watershed in Colorado, where diagnostics show that increasing temperatures are associated with decreasing streamflows. Using temperature information for lead years 2–6, two analyses are performed: (i) a retrospective hindcast for the climatological period (1981–2010) and (ii) a blind forecast for 2011–15. For the retrospective hindcast, including temperature information improved the percent error as compared with climatology. For the blind forecast, the multiyear temperature prediction for warming was skillful, but the corresponding multiyear average streamflow predictions from both approaches were counterintuitive: with the predicted warming, the multiyear average streamflow was predicted to be lower than the climatological mean; however, the observed multiyear average streamflow was higher than the climatological mean. This was due to above-average precipitation during the prediction time frame, particularly for one of the years. With that year removed, the multiyear streamflow average became lower than the climatological mean. Temperature provides a marginal source of streamflow predictability, but there will be substantial uncertainty until prediction skill for year-to-year climate variability, especially for precipitation, increases.

KEYWORDS: Streamflow; Climate prediction; Forecast verification/skill; Hydrologic models

1. Introduction

Multiyear climate prediction is a rapidly evolving field that aims to inform society of potential climate variability and change years to a decade in advance. Also called decadal prediction, outlooks at this time scale fill the gap between short-term (subseasonal to seasonal) and long-term (multidecadal to centennial) planning horizons. These decadal predictions are being informed by global climate models (GCMs) that are initialized based on current observations and run with projected anthropogenic forcing, thus offering potential to capture both the natural variability and the forced response (Meehl et al. 2009). The initialization makes these predictions distinct from their more well-known counterparts, climate change projections. GCM climate change projections are considered uninitialized, in that they are run continuously forward in time, being driven mostly by external forcing

(Goddard et al. 2012). Further, these predictions have been defined as distinct from “forecasts,” where forecasts imply that raw predictions are further processed to be disseminated as decision-making products (Goddard et al. 2012). Nevertheless, the multiyear to decadal time horizon has been identified as valuable to society (Vera et al. 2010), especially the water sector (Barsugli et al. 2009), which is already a sophisticated user of climate information (Raucher et al. 2015). As climate predictions on the time scale of years to a decade become more mainstream, guidance is needed to assist practitioners who wish to explore the utility of the predictions in this maturing field.

Multiyear predictions have become a coordinated effort for the climate community and are becoming increasingly accessible. Phase 5 of the Coupled Model Intercomparison Project (CMIP5) included decadal hindcast and prediction experiments (Taylor et al. 2012), and decadal prediction contributions are also planned for CMIP6 (Boer et al. 2016). There is also a large ensemble of decadal predictions and hindcasts using the Community Earth Systems Model (CESM; Yeager et al. 2018). Further, efforts to provide real-time experimental predictions are underway and regularly updated (Smith et al. 2013). From these and other efforts, decadal predictions have been shown to be skillful, particularly for temperature (Meehl et al. 2014; Smith et al. 2007; Yeager et al. 2018), though recent work has suggested that precipitation and circulation may be more skillful than previously believed (Smith et al. 2019).

A synthesis of current and emerging developments of subseasonal to decadal prediction was recently summarized by

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JAM-CD-20-0134.s1>.

^b ORCID: <https://orcid.org/0000-0002-1784-1346>.

Corresponding author: Erin Towler, towler@ucar.edu

DOI: 10.1175/JAMC-D-20-0134.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Merryfield et al. (2020), and there is an emerging case to make decadal predictions operational (Kushnir et al. 2019). Currently, the World Meteorological Organization (WMO) collects and provides experimental, real-time forecasts from global climate centers on the annual-to-decadal time scale (WMO 2020). Towler et al. (2018) develop a framework toward applying multiyear predictions, specifically showing how they can be presented in ways that are familiar to practitioners, that is, as discrete anomalies, like many climate change projections, or probabilistically, like seasonal climate forecasts. Towler et al. (2018) provide the initial steps toward the application of multiyear temperature predictions but stop short of delving into the additional considerations needed for their incorporation in impact models. This study picks up where Towler et al. (2018) left off, demonstrating the ways multiyear climate predictions can be incorporated into hydrologic modeling. Although there is literature on integrating subseasonal to seasonal forecasts (e.g., Baker et al. 2019; Towler et al. 2010) and climate change projections in water planning (e.g., Fowler et al. 2007), there is a paucity applying multiyear predictions into a specific impact-modeling application.

The contribution of this paper is to demonstrate the process and considerations of incorporating multiyear temperature predictions into impact modeling, as well as to discuss the advantages and limitations. Using an example from the water sector, multiyear temperature predictions are presented in two ways: discrete and probabilistic, which are examined using two common approaches to modeling hydrologic impacts, conceptual and empirical, respectively. Analyses are applied to both a retrospective hindcast for the climatological period (1981–2010) and a blind forecast for 2011–15. This is demonstrated by using 5-yr average temperature information for lead years 2–6 to predict streamflow in the upper Colorado River basin watershed in Colorado.

2. Case study

This analysis focuses on the upper Colorado River basin [UCRB; U.S. Geological Survey (USGS) hydrologic unit code (HUC) 140100] in Colorado. Much of Colorado's water supply comes from snowmelt runoff that originates in the mountains of Colorado, including the UCRB. Water from the UCRB is important for water resources for the state of Colorado and is also part of the headwaters that contribute to Colorado River inflows, which provides water supply to many western U.S. states, including by way of large reservoirs such as Lake Powell and Lake Mead.

Colorado River basin flows are sensitive to temperature and precipitation, although how sensitive has been subject to recent scientific debate. To investigate this, researchers have used different modeling approaches and assumptions, but have been able to compare results by normalizing in terms of temperature sensitivity, that is, percent change in annual flow per degree change in annual temperature. Udall and Overpeck (2017) show that flow declines have been driven mainly by increasing temperatures, with a sensitivity of $-7\% \text{ }^{\circ}\text{C}^{-1}$. However, Hoerling et al. (2019) argue that temperature alone accounts for one-third of the streamflow decline, or $-2.5\% \text{ }^{\circ}\text{C}^{-1}$, and

precipitation accounts for the remaining two-thirds of the decline. Milly and Dunne (2020) found discharge to be decreasing by $9.3\% \text{ }^{\circ}\text{C}^{-1}$. Although the present study focuses on only one subbasin of the Colorado River headwaters (the UCRB), these previous studies suggest that temperature predictions could be useful in understanding likely changes in UCRB streamflow. In this study, streamflow is examined at a location in the UCRB: the Colorado River at Cameo, Colorado (Fig. 1). Cameo is a critical management point on the system, as it is a senior water right for which upstream transbasin diversions can be curtailed when flows are insufficient (Yates et al. 2015). Annual diversions from the Colorado River into the South Platte and Arkansas River basins are about 300 000 acre feet (AF; 3.7 million m^3) on average, representing about 12% of the depleted flow at Cameo, with most diversions happening between 1 April and 1 October. Reservoir storage above Cameo is about 1 000 000 AF (1.23 million m^3) or less than one-third of the annual flows at Cameo (Yates et al. 2015).

a. Forecasting context

As indicated above, the UCRB is an important water source for the state of Colorado, as well as the Colorado River system that contributes to the water supply for many states in the western United States. Wood et al. (2020) discusses the state and details of streamflow forecasting in the Colorado River basin. In this paper, the incorporation of multiyear temperature predictions is explored in the context of general forecasting practices in the basin. On annual time scales, forecasters want to predict the streamflow volume for the coming water year. In the Colorado River basin, streamflow is dominated by spring and early summer snowmelt runoff. Predictions can begin in the preceding autumn, and, although in operational seasonal forecasting the predictions are updated monthly through winter and spring (adding snowpack information), here the focus is on the annual time scale, starting in October before there is appreciable snowpack. In the Colorado River basin, one way operational forecasters develop streamflow predictions is using ensembles; they run their hydrologic models forward in time using historical climate forcings from each year of the 30-yr climate normal period (1981–2010 as defined by the World Meteorological Organization). The resulting 30-member ensemble is comprised of water-year streamflow volumes and represents a climatological or baseline risk profile that can be used for planning. For instance, the ensemble average, median, and/or upper and lower quantiles can be used to drive management or decision models, or the profile can be used to estimate the likelihood of exceeding a particular management trigger.

This general ensemble forecasting context is used as a starting point for incorporating multiyear temperature predictions, but the baseline ensemble needs to be perturbed to reflect the multiyear temperature prediction. In this study, the focus is on a 5-yr prediction period of lead years 2–6. That is to say, a prediction issued in November of 1981 is used to predict water years 1983–1987, and so on. Technically, this is a lead of 11 months (rather than 12 months), but this

FIG. 1. Location of the Cameo gauge (red star) on the Colorado River in the UCRB (outlined in black), Colorado; the red lines represent transbasin diversion structures, and the yellow circles are the main population centers.

was necessary since the predictions are run annually on 1 November; this is further discussed in [section 3a\(1\)](#) below.

b. Diagnostics

Streamflow data for the Colorado River at Cameo are from the USGS gauge (09095500 Colorado River near Cameo). Located on the western slope of Colorado, runoff volumes are dominated by snowmelt runoff in the UCRB. Temperature data over the UCRB domain were also examined from the Daymet observational dataset ([Thornton et al. 2016](#); <https://daymet.ornl.gov/getdata>), which has a 1-km resolution. Two-meter average temperature (T_{avg}) was calculated as a function of maximum temperature (T_{max}) and minimum temperature (T_{min}), such that $T_{avg} = 0.6 \times T_{max} + 0.4 \times T_{min}$ ([Thornton et al. 1997](#)).

The strength of the relationship between water-year streamflow at Cameo and UCRB temperature is examined. [Figure 2a](#) shows the scatterplot between the water-year volume anomalies and the water-year average temperature over the UCRB watershed. [Figure 2a](#) shows that for Cameo the relationship from 1981 to 2010 (black dots) is negative with a correlation coefficient r of -0.58 , which decreases slightly using the period from 1981 to 2015 (-0.50 ; 2011–15 is in blue).

Although this study only demonstrates incorporating multiyear temperature predictions, it is important to also examine the relationship between precipitation and streamflow in the UCRB. [Figure 2b](#) shows that the relationship between precipitation and streamflow is positive, and stronger than

the relationship between flow and temperature (i.e., $r = 0.86$ for the 1981–2011 period), a point that will be revisited in the discussion section.

3. Method

The method is organized in two steps: first, the acquisition of the multiyear temperature predictions, and second, the generation of the streamflow prediction ensemble that is conditioned on the temperature predictions.

a. Step 1: Acquire multiyear temperature prediction

Although multiyear predictions are still considered to be experimental, temperature predictions on multiannual time scales are becoming more accessible to researchers and practitioners. [Towler et al. \(2018\)](#) show how multiyear temperature predictions can be presented as a deterministic anomaly or as a tercile-based probability. The deterministic anomaly approach is how climate change projections are often communicated, whereby relative changes are often presented as a deterministic anomaly, or delta, relative to a historical baseline period (e.g., 30-yr climatology). Deltas can be calculated at every point in a gridded dataset (at the resolution of the underlying climate model) or aggregated to some larger domain. For instance, water managers are interested in watershed scales, and deltas can be calculated for a particular watershed of interest (e.g., [Towler et al. 2018](#)). This is similar to a product developed by [Baker et al. \(2019\)](#) that presents real-time subseasonal to

FIG. 2. Scatterplot between Colorado River at Cameo streamflow (acre feet and kilometers cubed) anomalies and water-year (left) temperature and (right) precipitation averaged over the UCRB. Black dots are 1981–2010, and blue labeled dots are 2011–15; lines are linear fit. The “e” configuration on the y axis indicates that the first digit should be multiplied by 10 raised to the power of the last digit following the e.

seasonal forecasts on watershed scales across the United States.

Another familiar approach is the use of tercile-based probabilities. Operational seasonal forecasts issued by NOAA’s Climate Prediction Center and Columbia University’s International Research Institute for Climate and Society (IRI) are presented probabilistically (Mason and Goddard 2001), whereby probabilities are shown for the likelihood of being below, near, or above the climatological terciles. Again, these can be calculated for the spatial scale of interest, such as watersheds (e.g., Towler et al. 2018). Offering decadal predictions in both of these formats may help to bridge with more potential users, as it has been shown that users prefer familiar formats (Taylor et al. 2015).

Next, the climate model data are briefly described, as well as how the delta anomalies and terciles are calculated. However, it is noted that, in an operational context, multiyear predictions would likely be products provided by an operational agency.

1) MULTIYEAR TEMPERATURE PREDICTIONS

The large ensemble of multiyear decadal predictions [Decadal Prediction Large Ensemble (DPLE)] was taken from the National Center for Atmospheric Research’s CESM, for which each annual prediction includes 40 members (Yeager et al. 2018). Hindcasts are from annual initializations from 1 November 1981 through 2010 (for a total of 30 start dates). The CESM decadal predictions are drift corrected using the CLIVAR protocol (International CLIVAR Project Office 2011), as was done in Towler et al. (2018) and other studies (e.g., Meehl and Teng 2012, 2014a,b). As described in Meehl et al. (2014), drift correction is required because model runs that are initialized to observations “drift” back to their preferred state. Further, the drift is time dependent (more rapid at first), and is therefore distinct, and more complicated to correct

for, than bias correction techniques typically applied for climate change projections.

From the DPLE, 2-m T_{max} and T_{min} are examined; as was done for the observations, average temperature is calculated as $0.4 \times T_{\min} + 0.6 \times T_{\max}$ (Thornton et al. 1997). The DPLE is drift corrected to temperature observations; in this case they were drift corrected to the PRISM Gridded Climate Data Group dataset (http://www.prism.oregonstate.edu/documents/PRISM_downloads_FTP.pdf), which has a 4-km resolution.

Towler et al. (2018) shows that there is skill in the temperature hindcasts for the UCRB using a previous version of the CESM, that is, CCSM4, which only had 10-members for each initialization. Yeager et al. (2018) shows that the DPLE from CESM shows skill improvements over the CCSM4. Here, several skill analyses were performed on the DPLE to show its strengths and value as it relates to this study. Figure 3 shows a spatial comparison over the UCRB of the temperature and precipitation hindcasts for lead years 2–6: this shows high anomaly correlation coefficient (ACC) values for temperature but lower and more variable values for precipitation. This is consistent with other studies and is why the focus here is on using temperature. Second, to show how predictability varies with period averaging, Fig. 4 shows the ACC for temperature using the DPLE over the UCRB for different multiyear average periods. In general, it can be seen that the ACC values are lowest for the individual water years (2–2, 3–3, 4–4, etc.), and increase with period averaging. This is important to note because although potential users of decadal information might want predictions that map to the sequencing of each individual year, there is less predictability. This presents a potential mismatch between what users want and what science can provide, and this point is further addressed in the discussion section. Finally, to quantify the value of using an initialized dataset, the DPLE can be compared with a corresponding large

FIG. 3. ACC skill assessment plot for lead years 2–6 using the DPLE hindcasts: (left) Tavg ACC and (right) precipitation ACC over the UCRB.

ensemble (LE) from CESM (Kay et al. 2015), which uses the exact same model configuration and forcings but is not initialized. For multiyear periods starting at lead year 2, Fig. 5 shows that the DPLE has consistently higher ACC values than the LE, which indicates that the initialization is offering some value.

For this paper, using temperature information for lead years 2–6, two analyses are performed: (i) a retrospective hindcast for the climatological period (1981–2010), in which 24 prediction periods were evaluated (i.e., the first being water years 1983–1987, and the last being water years 2016–10, and (ii) a blind forecast for the 2011–15 period. The latter is the same prediction period explored in Towler et al. (2018).

2) DISCRETE DELTA (ANOMALY) APPROACH

The distribution of the predicted temperature deltas (anomalies) is calculated for the prediction period of interest as compared to the hindcasts from the climatological period, defined as water years 1981–2010. The delta distribution is examined for the 200 (= 5 years $\times$ 40 members) annual temperatures for the UCRB. These values are what were used

in the skill assessments described in the previous section [3a(1)] and are examined for blind forecast; here the November 2009 hindcast is used to obtain the 5-yr period for lead years 2–6 (October 2010–September 2015), with each year having 40 members. Table 1 shows the select percentiles from the delta distribution. The 50th percentile (Q50) of the distribution is 0.4°C, and the 70th percentile (Q70) was 0.9°C, which was close to the observed. The 30th percentile (Q30) of the distribution was 0.2°C.

For the retrospective hindcast, the observed deltas for each of the 24 prediction periods were calculated. This provides an estimate of the “perfect forecast” and offers insight to the limits of predictability of using the delta approach. It should be pointed out that the delta could be negative; this was seen in the earlier prediction periods (e.g., the November 1981 forecast for water years 1983–87 had a delta of -1.1°).

3) TERCILE-BASED APPROACH

Terciles divide a variable into three equal categories with respect to climatology. Values placed in each of the tercile categories can be referred to as below normal, near normal,

FIG. 4. ACC for Tavg using the DPLE hindcasts over the UCRB for different averaged water-year forecast lead periods. The period of 2–6 is used in this study.

FIG. 5. ACC for Tavg using the DPLE hindcasts and uninitialized LE over the UCRB for averaged water-year forecast lead periods. The period of 2–6 is used in this study.

and above normal. Because a 30-yr climatological period (1981–2010) is being used, it follows that each tercile category contains 10 years. In the blind forecast, the November 2009 hindcast for the 2011–15 period has 200 annual temperatures for the UCRB; here 157 of the 200 members are in the above-normal category (78%). Table 1 shows the tercile probabilities for 2011–15 from the DPLE (i.e., 78% from the above normal, 3% from the near normal, and 19% from the below normal), as well as from what was observed (i.e., 80% above normal, 20% near normal, and 0% below normal). For the retrospective hindcast, temperature terciles for each 5-yr hindcast period, from both the DPLE and from what was observed (i.e., a perfect forecast) are also calculated.

b. Step 2: Generate streamflow predictions conditioned on temperature

Once the temperature prediction is obtained (e.g., Table 1), to be useful for water resources planning it needs to be used in conjunction with an impact model that can translate it to streamflow. In this study, streamflow is generated using two approaches to a hydrological model. Using a simple dichotomy introduced by Clarke (1973), the hydrological model can be conceptual or empirical, where conceptual models aim to incorporate physical understanding of the process. Here, one of each type of model is described, and how they could be used with the obtained decadal forecasts from step 1.

1) CONCEPTUAL HYDROLOGICAL MODEL

There are a broad range of conceptual hydrological models with varying levels of complexity that have been developed (e.g., Kampf and Burges 2007; Clark et al. 2011). Any conceptual model that is forced with climate data can be used with multiyear temperature predictions. There are a variety of ways in which forcing data can be perturbed or modified to reflect climate model output (see, e.g., Prudhomme et al. 2002; Fowler et al. 2007), including the pseudo-global warming approach (Rasmussen et al. 2011), which, when applied only to temperature, is called the delta approach. In this study, the delta approach is applied for its simplicity, as well as for its congruence

TABLE 1. Observed and predicted average temperature discrete deltas (anomalies) and probabilistic tercile values between climatology (1981–2010) and 2011–15 over the UCRB watershed.

	Obs	DPLE ^a
Anomaly (°C)	0.92	0.4 (0.2–0.9) ^b
Probability (%)		
Below normal	0	19
Near normal	20	3
Above normal	80	78

^a Initialized November 2009 for October 2010–September 2015.

^b Q50 (Q30–Q70).

with the anomalies that are commonly calculated from global models in climate change studies. Hence it is fitting that it could also be used with multiyear predictions, which share many characteristics with multidecadal to centennial projections. Although this type of period averaging does not offer insight on the year-to-year variability, it is still useful in watersheds that contribute to large reservoirs with year-over-year storage, such as how UCRB streamflow contributes to Lake Powell and Lake Mead.

The conceptual model that is used is the Water Evaluation and Planning (WEAP; Yates et al. 2005), which has been tailored to the UCRB (USGS HUC 140100), as described in Yates et al. (2015). WEAP operates at the catchment level and includes a physically based hydrologic simulation capability that is driven by climate forcing. Climate forcing input for WEAP was developed from the Daymet dataset (Thornton et al. 2016), where average temperature and precipitation for each catchment within the UCRB are used to drive the hydrologic simulation using a weekly time step. Here the focus is on changing the climate forcings (e.g., Yates et al. 2009) to be consistent with the multiyear temperature predictions obtained in Step 1.

The UCRB WEAP model includes water demands and infrastructure (reservoirs, diversions, etc.), and incorporates the management rules and water rights that determine the trans-basin diversion timing and volume (Yates et al. 2015). For example, Cameo is a senior water right that “calls out” upstream, junior water rights, with the model mimicking these calling procedures, and thus the decadal predictions being made represent managed conditions.

For the retrospective hindcast, to create a climatological WEAP model baseline, for each prediction period, the five years being predicted are dropped, and the remaining individual water years or traces, are run through the WEAP model, creating modeled streamflow for each year. For the blind forecast, all 30 individual water years (1981–2010) or traces, can be run through the WEAP model. These can be compared to the observations to indicate how well the conceptual model performs, but more importantly, it creates a modeled climatological baseline to which relative changes from adding a temperature delta can be estimated. The observed temperature and precipitation of the five individual water years (2011–15), are run through the WEAP model, to compare with the prediction.

To generate streamflow predictions conditioned on the multiyear temperature predictions, the same input traces from the climatological baseline are run through WEAP, except now the temperature delta for each prediction period is added uniformly to the climatological temperature forcings. For the blind forecast, two delta scenarios are run: the Q30 and Q70 deltas from Table 1 (0.2° and 0.9°C) are added to each weekly temperature of 1981–2010. For the retrospective hindcast, the observed deltas from each prediction period are used (i.e., a perfect forecast), to provide the limits of predictability of the delta approach. Precipitation stays the same, even though this could create some physical inconsistencies; this is addressed further in the discussion section.

2) EMPIRICAL HYDROLOGIC MODEL

Empirical, or statistical models are a viable alternative to using conceptual hydrologic models. In this paper, WEAP had been developed and calibrated for the UCRB, but often, this is not the case and building a model for a new catchment can be data and resource intensive, which is not always feasible. There are many statistical models that can be used to simulate streamflow (Salas 1993), ranging from simple to complex. In this paper, a simple empirical approach based on conditional resampling is presented.

Recent work has investigated the relationship between temperature and streamflow in the Colorado River basin (Udall and Overpeck 2017; Hoerling et al. 2019; Milly and Dunne 2020), but when developing an empirical model, it is best if data can be examined for the location of interest. As mentioned in section 2, the Cameo location is the focus, for which the relationship between temperature and streamflow has been shown (Fig. 2a). Given that the water-year temperature predictions are skillful in the UCRB (Fig. 3), the temperature predictions are used to conditionally resample the associated streamflow.

Specifically, the historical average water-year temperatures are ranked in ascending order, with the bottom third partitioned as the below-normal years, the middle third as the near-normal years, and the top third as the above-normal years. The corresponding streamflows for each water year are designated to each bin; for the 1981–2010 climatological period there are 10 streamflow volumes in each bin. For Cameo (Fig. 6), the streamflow medians increase with decreasing temperature bins: that is, the below-normal temperature bin has the highest streamflow median, and the above-normal temperature bin has the lowest streamflow median. However, there is notable overlap in the interquartile range (IQR; i.e., the 25th–75th percentiles) of the distributions.

Next, the historical streamflow volumes are bootstrapped (Efron and Tibshirani 1993) with replacement. For the retrospective hindcast, the 5 years being predicted are dropped from the climatology and resampling ensemble. This was done using the actual DPLE hindcast predictions, as well as by using the observed terciles (i.e., a perfect forecast). For the blind forecast, the bin sampling uses the full climatological record, in accordance with the probabilities in Table 1. This creates 100 streamflow volume ensemble members that reflect the multi-year temperature prediction.

FIG. 6. Water-year streamflow volumes (KAF) for Cameo binned by UCRB temperature tercile bin for 1981–2010.

4. Results

a. WEAP conceptual model results

First, the results of the blind forecast are examined. To start, WEAP is run for each year (trace) of the climatological period 1981–2010 resulting in 30 individual traces. For Cameo, using the observed forcings, WEAP is able to adequately capture the observed water-year streamflow volumes (Fig. 7). The WEAP median [2230 thousand AF (KAF)] is slightly lower than the observed median volume (2582 KAF). This brings up the first consideration of using a conceptual model, which is that while hydrologic models are representations of reality, they have inherent biases. This leaves the user with several choices: to bias correct the WEAP model output or to look at the relative changes between the baseline and the delta scenarios. In this work, the latter is adopted for running the delta scenarios, as indicated in the next paragraph.

Next, the same 30 traces are run, except the forcing temperature inputs are perturbed by uniformly applying the temperature delta, and precipitation remains unchanged. Although this is a discrete anomaly, a sense of the range can be seen by using the lower and upper deltas from Table 1 (0.2° and 0.9°C). Figure 7 shows that adding the delta shifts the median down, first by -1.7% (for 0.2°C) and -5.1% (for 0.9°C); this translates to a reduction between 5.6% and 8.5% $^{\circ}\text{C}^{-1}$, which is in line with the aforementioned studies. Because the delta is applied to all 30 traces, the flow distribution remains similar, but shifted toward lower values.

For the retrospective hindcast, WEAP is run using each year (trace) that is not in the prediction period (those five years are dropped) to develop a streamflow climatology ensemble (WEAP_clim), and then using the observed deltas (perfect forecast) for each prediction period (WEAP_delta). Percent errors are calculated from the median of the five years dropped versus WEAP_clim and WEAP_delta. From Fig. 8 (left), the median percent error decreases from 9.8% to 4.3% if the deltas are utilized rather than climatology. Positive percent errors mean that WEAP underestimated the streamflow 5-yr

FIG. 7. For the blind forecast, climatological-period (1981–2010) water-year volumes from the observations, WEAP simulation, and WEAP simulation plus 0.2° and 0.9°C delta. The horizontal line in the box is the median, with the median value given above it; the percent change from the WEAP baseline is in parentheses.

average, and negative percent errors mean WEAP overestimated the streamflow 5-yr average. In Fig. 8 (middle and right), the percent errors are broken out by if the 5-yr period being predicted was observed to be warm versus cool or wet versus dry: in terms of the medians, using the delta approach decreased the percent error in normal temperature periods and for precipitation periods that were normal or wet, as compared to using the baseline climatology.

b. Resample empirical model results

For the blind forecast, the observed water-year streamflows are resampled in accordance with the percentages in Table 1, resulting in a 100-member statistical simulation. Since these are resampled from observations, the results can be directly compared with the observed climatology. Figure 9 shows that the resample, strongly tilted toward the above-average temperature bin (i.e., 78% from Table 1), results in a lower median than the baseline, similar to the WEAP delta temperature results. The resample results in a median flow decrease of 9.7%, which is higher than found in the WEAP scenarios. It is also of note that the IQR is smaller than the observations, although a similar range is represented. In addition, the perfect forecast resample was examined (i.e., from Table 1: 80% from above normal, 20% from near normal, and 0% from below normal), and the results were quite similar (results not shown). The one difference was that it resulted in a smaller range of streamflows in terms of the maximum, since there were no streamflows resampled from the below-average bin.

For the retrospective hindcast, the statistical resample is done using the actual predictions from the DPLE, as well as using the perfect forecast of the observed terciles. It was straightforward to do both for the statistical resample, since this approach requires very little computational resources. For both, the five years in the prediction period are dropped. Results were similar, and only the actual DPLE results are discussed here. Figure 10 (left) shows that the absolute value of the median percent error decreases from −5.7% to 3.1% using the weighted resample. In Fig. 10 (middle and right), the percent errors are broken out by if the 5-yr period being predicted was warm versus cool or wet versus dry: the resample showed

FIG. 8. For the retrospective hindcast, the (left) percent errors using the WEAP climatology (Clim) and using the WEAP climatology with the deltas (Delta), along with the percent errors broken out by (center) cool vs warm and (right) dry vs wet prediction periods.

FIG. 9. For the blind forecast, water-year volumes from observations from the climatological period (1981–2010), the weighted resample, and observations from 2011 to 2015 and 2012 to 2015. The horizontal line in the box is the median, with the median value given above it; the percent change from climatology is in parentheses.

improvements in median percent errors when the 5-yr period was warm or dry.

c. Observed versus prediction results

For the blind forecast, what was predicted versus what was observed in 2011–15 is examined. From Table 1, the 2011–15

period had an average temperature increase of 0.9°C relative to the 1981–2010 baseline period, which was consistent with the higher end (Q70) of the DPLE predicted range, and one of the deltas used in the WEAP blind forecast. Further, the above-average tercile prediction and observation was very close (80% vs 78%, Table 1). However, examining the 2011–15 years compared to the 1981–2010 climatology, the observed median flow is higher by 18% (Fig. 9, Table 2) and the observed average flow is higher by 4.0% (Table 2). From Table 2, the ensemble average predicted flow is -13% from the resample and -5.0% from WEAP 0.9 delta scenario. These results seem counterintuitive to the expectation that flows would decrease with warming in the 2011–15. Similar results are seen when the WEAP model is driven with the 2011–15 observed temperature and precipitation (Fig. S1 in the online supplemental material).

Table 3 helps to disentangle this by looking year by year. In Table 3, the streamflows at Cameo from 2011 to 2015 are listed in decreasing flow order, along with their associated precipitation and temperature anomalies. As precipitation decreases, flow also decreases monotonically. There is less of a discernable pattern to the temperature anomalies and terciles. Temperature is above normal in four of the five years, and precipitation is above normal in three of five years. Flows end up being below normal when temperature is above normal, and precipitation is either normal or below normal. If 2011 is removed, which had above-average precipitation (102 mm yr^{-1} anomaly), then 2012–15 flow average fits more with the resample and WEAP ensemble; the percent change from 1981 to 2010 is -8.2% (median) and -12% (average), which are more in line with the WEAP $+0.9$ delta scenario and the resampling ensemble.

FIG. 10. For the retrospective hindcast, (left) percent errors using the climatology (from observations) and using the resample based on the DPLE. Also shown are percent errors broken out by (center) cool vs warm and (right) dry vs wet prediction periods.

TABLE 2. Median and average streamflow values at Cameo, as well as percent changes from 1981 to 2010 to (i) observations from 2011 to 2015, (ii) observations from 2012 to 2015, (iii) the weighted resample, and (iv) the WEAP delta method.

	Observations			Resample	WEAP	
	1981–2010	2011–15	2012–15		1981–2010	+0.9
Median (KAF)	2581.83	3059.25	2368.86	2331.97	2229.62	2115.96
Percent change		18%	–8.2%	–9.7%		–5.1%
Avg (KAF)	2782.89	2895.19	2447.70	2412.40	2547.35	2419.51
% change		4.0%	–12%	–13%		–5.0%

The retrospective hindcasts can also be examined to shed some light on these results. The 2011–15 prediction period was a combination of warm and wet; interestingly the weighted resample was better than climatology when it was warm and dry (Fig. 10), and the WEAP model did better than climatology when the prediction period was normal temperature and precipitation was normal or wet (Fig. 8). This shows how the retrospective hindcast can be valuable for illustrating climate conditions in which each of the models offered advantages and disadvantages over a baseline climatology.

5. Discussion

a. Advantages and limitations of multiyear temperature predictions

This paper demonstrates the use of multiyear average temperature predictions being used to predict multiyear average streamflows. This raises several points: first, even if the multiyear temperature prediction is skillful, there can be year-to-year variability that may result in a counterintuitive prediction of the multiyear streamflow average. For instance, even though the multiyear temperature prediction captured the observed average temperature (0.9°C) and probability of being in the above-average tercile (78%, see Table 1), both the conceptual-anomaly and empirical-probabilistic ensemble averages underestimate the observed 2011–15 streamflow average. This was mainly due to 2011, which despite having above-average temperature, also had above-average precipitation. For the conceptual anomaly approach, the ensemble's average streamflow represents the sensitivity to the temperature delta given average precipitation. Similarly, for the empirical probabilistic approach, the ensemble's average streamflow represents the expected outcome given the weighted average precipitation, that is, weighted by the probabilities in Table 1.

Hence the predicted streamflow ensemble average is the expected value based on average precipitation or weighted average precipitation. This is an important caveat for water managers exploring the use of multiyear temperature predictions since streamflow is driven by both evapotranspiration and precipitation. While temperature can be used as a proxy for evapotranspiration, not including precipitation reduces the potential predictability. The marginal predictability gained from temperature versus precipitation will depend on time of year and location, similar to streamflow forecasting work that has shown when and where skill can be gained from initial conditions versus climate predictions (Wood et al. 2016). Although multiyear precipitation has shown encouraging improvements (Smith et al. 2019), an intermediate approach would be to incorporate precipitation scenarios with high/normal/low precipitation change factors (see Mauran et al. 2010). In locations where streamflow is inversely influenced by temperatures, such as the UCRB, managers need to be aware that given increasing temperature trends, streamflow predictions only using temperature predictions will be conservative, in that they assume average precipitation, and have considerable uncertainty since they do not account for precipitation variability or change.

Note that average multiyear flow predictions are most useful in basins where the streamflow contributes to large reservoirs with multiyear storage, such as how UCRB streamflow contributes to the downstream reservoirs of Lakes Powell and Mead. However, water managers are also interested in year-to-year variability since the order of years matters for many management decisions on this time scale. Nevertheless, the individual year temperature predictions are not yet skillful (Fig. 4), and credibility is key to stakeholder uptake (Cash et al. 2003). This information can provide feedback to the iterative process of continually adjusting scientific research agendas and user expectations; this type of coproduction of knowledge has

TABLE 3. Observed streamflow volumes at Cameo from 2011 to 2015 listed in descending order, along with associated precipitation and temperature anomalies over the UCRB watershed.

Water year	Streamflow vol (KAF)	Precipitation		Temperature	
		Tercile	anomaly (mm yr ⁻¹)	Tercile	anomaly (°C)
2011	4685	Above normal	102	Above normal	0.6
2014	3403	Above normal	74	Above normal	0.3
2015	3059	Near normal	55	Above normal	1.5
2013	1678	Below normal	5	Near normal	0.6
2012	1650	Below normal	–151	Below normal	1.6
Avg	2895		17		0.9

been shown to be effective at increasing useability (Dilling and Lemos 2011).

b. Conceptual versus empirical for incorporating multiyear predictions

In this paper, WEAP is used for the conceptual model. This was possible because a WEAP model had already been developed and calibrated for this basin, but this is not always the case. As a conceptual model, WEAP can be used to investigate processes contributing to the changes in streamflow (e.g., an increase in evapotranspiration and decrease in soil moisture), which can help provide process-based understanding to changes. Explicitly investigating the underlying physical processes is not possible with the data-driven empirical approach. A big advantage of WEAP is that it includes management, which can be an important factor in hydrologic systems that are increasingly managed. This paper examined streamflows on the Colorado River at Cameo, which is a critical management point on the system, but the model can provide streamflows at any point in the system, including ungauged basins, and incorporate management triggers such as reservoir storage/releases to support decision-making (Yates et al. 2015). If and when multiyear precipitation predictions become skillful, the conceptual model can readily ingest these in the form of precipitation change factors. Further, bias-corrected multiyear temperature and precipitation could be run directly in the model, as has been done in climate change studies and products (see, e.g., Bureau of Reclamation's Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections website: https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html#About), or seasonal deltas could be applied. Finally, the WEAP model could also be driven with probabilistic tercile values from Table 1: for instance, resampling 78% of above-average temperature years, 3% near-normal years, and 19% below-normal years, and running those through WEAP would also create a weighted streamflow ensemble. However, an advantage of developing a more complex, process-based conceptual model is the ability to see how increasing temperatures influence streamflows in a physically credible way.

The demonstrated empirical approach is a statistical resampling scheme that is informed by the tercile probabilities from the multiyear predictions. As mentioned, conditioning on precipitation would be advantageous if it were to become more skillful, and methods exist to resample based on both temperature and precipitation (see Briggs and Wilks 1996). Further, there are statistical resampling approaches that can be used with the anomaly, such as nearest neighbor bootstrapping (e.g., Lall and Sharma 1996; Baker 2019). One issue with resampling based on temperature is that temperature has an increasing trend, and as was seen, four of five years (80%) were in the above-average tercile, which limits the resampling of the below-normal and near-normal temperature years. This could result in an underdispersed prediction, that is, where there is not enough variability to capture the full range of possible responses. The resample approach would also be useful to try with a longer record, for example paleoclimate (Brekke 2009), and would also benefit from continually updating the climatological period iteratively every year. Last, developing a

predictive statistical model that could be used with the delta anomaly is possible—for instance, an empirical transfer function between temperature and streamflow—but the disadvantage is that to use the anomaly, the transfer function relies on extrapolation, and has less physical basis than can be explored in a conceptual model.

6. Conclusions

This study demonstrates approaches to incorporating 5-yr average temperature information for lead years 2–6 into hydrological simulation. A variety of analyses were performed, including a retrospective hindcast and a blind forecast, as well as the pairing of a discrete delta with a conceptual hydrological model and probabilistic tercile likelihoods with a statistical resample method. The focus is on temperature predictions because they are skillful and influence streamflow in the basin, where increasing temperatures are associated with lower streamflows. The retrospective hindcast showed that including temperature information can improve the percent error as compared to climatology. For the blind forecast, when positive deltas are added to the baseline meteorological forcings, the WEAP conceptual model simulates lower streamflow ensemble averages. Similarly, the resample streamflow ensemble average shifts toward lower streamflows when weighted toward higher temperatures. Despite observed temperatures being consistent with the multiyear prediction, the average streamflow prediction for years 2–6 was counterintuitive: the 2011–15 average observed streamflow was higher than the 1981–2010 average observed streamflow. This was mainly due to 2011, which, despite experiencing above-average temperature, also had above-average precipitation. Temperature affects evapotranspiration, which influences streamflows, but does not account for moisture inputs from precipitation, reducing the potential predictability.

Many water managers are already sophisticated users of climate information, and multiyear temperature predictions would be a logical next step to add to their climate portfolio. Many already have hydrologic modeling capabilities, either conceptual models that can readily ingest multiyear climate information to simulate streamflow or monitored streamflow data that they can strategically condition using a simple empirical approach. Conditioning streamflows on temperatures provides a marginal source of predictability and can be used by managers as conservative scenarios to explore how streamflows in the basin may shift with respect to temperature. Critically, managers need to acknowledge that the prediction does not account for precipitation, which is a main source of streamflow predictability. Until multiyear precipitation skill increases, precipitation scenarios could be used to quantify system sensitivity. Further, this study did not account for seasonal variability, even though warm season temperatures have been shown to be influential in reducing streamflows in the Colorado River (Das et al. 2011). The DPLe does show some variability in predictability by season (Fig. S2 in the online supplemental material); future studies could explore how varying the delta or terciles by season affects the results.

Experimental, real-time forecasts from global climate centers on annual to decadal time scales are already being provided (WMO), and there is momentum toward making decadal predictions operational (Kushnir et al. 2019). As such, this paper offers insight and a first step toward how they could potentially be applied to hydrologic simulation.

Acknowledgments. This material is based upon work supported by the National Center for Atmospheric Research (NCAR); NCAR is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement 1852977. This work was partially funded by NSF Grant AGS-1419563 as part of the Understanding Decision–Climate Interactions on Decadal Scales (UDECIIDE) project. David Yates acknowledges the OAR Climate Program Office (CPO) project NA16OAR4310135. We thank Ming Ge (NCAR) for providing data and the spatial skill map.

Data availability statement. The CESM-DPLE data can be accessed online (<https://www.cesm.ucar.edu/projects/community-projects/DPLE/>).

REFERENCES

- Baker, S. A., 2019: Development of sub-seasonal to seasonal watershed-scale hydroclimate forecast techniques to support water management. Ph.D. dissertation, University of Colorado, 168 pp., <https://search.proquest.com/openview/86480abe8a4f1b7c3f0bccc9bf5142ac/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- , A. W. Wood, and B. Rajagopalan, 2019: Developing sub-seasonal to seasonal climate forecast products for hydrology and water management. *J. Amer. Water Resour. Assoc.*, **55**, 1024–1037, <https://doi.org/10.1111/1752-1688.12746>.
- Barsugli, J., C. Anderson, J. B. Smith, and J. M. Vogel, 2009: Options for improving climate modeling to assist water utility planning for climate change. Western Utilities Climate Alliance White Paper, 144 pp., <https://www.wucaonline.org/assets/pdf/pubs-whitepaper-120909.pdf>.
- Boer, G. J., and Coauthors, 2016: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3751–3777, <https://doi.org/10.5194/gmd-9-3751-2016>.
- Brekke, L. D., 2009: Long-term planning hydrology based on various blends of instrumental records, paleoclimate, and projected climate information. U.S. Bureau of Reclamation Rep., 123 pp., https://www.usbr.gov/research/projects/download_product.cfm?id=2553.
- Briggs, W. M., and D. S. Wilks, 1996: Extension of the climate prediction center long-lead temperature and precipitation outlooks to general weather statistics. *J. Climate*, **9**, 3496–3504, [https://doi.org/10.1175/1520-0442\(1996\)009<3496:EOTCPC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<3496:EOTCPC>2.0.CO;2).
- Cash, D. W., W. C. Clark, F. Alcock, N. M. Dickson, N. Eckley, D. H. Guston, J. Jäger, and R. B. Mitchell, 2003: Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci. USA*, **100**, 8086–8091, <https://doi.org/10.1073/pnas.1231332100>.
- Clark, M., D. Kavetski, and F. Fenicia, 2011: Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resour. Res.*, **47**, 3901, <https://doi.org/10.1029/2010WR009827>.
- Clarke, R. T., 1973: A review of some mathematical models used in hydrology, with observations on their calibration and use. *J. Hydrol.*, **19** (1), 1–20, [https://doi.org/10.1016/0022-1694\(73\)90089-9](https://doi.org/10.1016/0022-1694(73)90089-9).
- Das, T., D. W. Pierce, D. R. Cayan, J. A. Vano, and D. P. Lettenmaier, 2011: The importance of warm season warming to western U.S. streamflow changes. *Geophys. Res. Lett.*, **38**, L23403, <https://doi.org/10.1029/2011GL049660>.
- Dilling, L., and M. C. Lemos, 2011: Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Global Environ. Change*, **21**, 680–689, <https://doi.org/10.1016/j.gloenvcha.2010.11.006>.
- Efron, B., and R. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- Fowler, H., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578, <https://doi.org/10.1002/joc.1556>.
- Goddard, L., and Coauthors, 2012: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, <https://doi.org/10.1007/S00382-012-1481-2>.
- Hoerling, M., J. Barsugli, B. Livneh, J. Eischeid, X. Quan, and A. Badger, 2019: Causes for the century-long decline in Colorado river flow. *J. Climate*, **32**, 8181–8203, <https://doi.org/10.1175/JCLI-D-19-0207.1>.
- International CLIVAR Project Office, 2011: Data and bias correction for decadal climate predictions. CLIVAR Publ. Series 150, 3 pp., https://eprints.soton.ac.uk/171975/1/ICPO150_Bias.pdf.
- Kampf, S. K., and S. J. Burges, 2007: A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water Resour. Res.*, **43**, W05423, <https://doi.org/10.1029/2006WR005370>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Kushnir, Y., and Coauthors, 2019: Towards operational predictions of the near-term climate. *Nat. Climate Change*, **9**, 94–101, <https://doi.org/10.1038/s41558-018-0359-7>.
- Lall, U., and A. Sharma, 1996: A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resour. Res.*, **32**, 679–693, <https://doi.org/10.1029/95WR02966>.
- Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, <https://doi.org/10.1029/2009RG000314>.
- Mason, S. J., and L. Goddard, 2001: Probabilistic precipitation anomalies associated with ENSO. *Bull. Amer. Meteor. Soc.*, **82**, 619–638, [https://doi.org/10.1175/1520-0477\(2001\)082<0619:PPAAWE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2).
- Meehl, G. A., and H. Teng, 2012: Case studies for initialized decadal hindcasts and predictions for the Pacific region. *Geophys. Res. Lett.*, **39**, L22705, <https://doi.org/10.1029/2012GL053423>.
- , and —, 2014a: CMIP5 multi-model hindcasts for the mid-1970s shift and early 2000s hiatus and predictions for 2016–2035. *Geophys. Res. Lett.*, **41**, 1711–1716, <https://doi.org/10.1002/2014GL059256>.
- , and —, 2014b: Regional precipitation simulations for the mid-1970s shift and early-2000s hiatus. *Geophys. Res. Lett.*, **41**, 7658–7665, <https://doi.org/10.1002/2014GL061778>.
- , and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1486, <https://doi.org/10.1175/2009BAMS2778.1>.
- , and Coauthors, 2014: Decadal climate prediction: An update from the trenches. *Bull. Amer. Meteor. Soc.*, **95**, 243–267, <https://doi.org/10.1175/BAMS-D-12-00241.1>.

- Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, 1252–1255, <https://doi.org/10.1175/BAMS-D-19-0037.1>.
- Milly, P. C. D., and K. A. Dunne, 2020: Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, **367**, E869–E896, <https://doi.org/10.1126/science.aay9187>.
- Prudhomme, C., N. Reynard, and S. Crooks, 2002: Downscaling of global climate models for flood frequency analysis: Where are we now? *Hydrol. Processes*, **16**, 1137–1150, <https://doi.org/10.1002/hyp.1054>.
- Rasmussen, R., and Coauthors, 2011: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *J. Climate*, **24**, 3015–3048, <https://doi.org/10.1175/2010JCLI3985.1>.
- Raucher, K., R. Raucher, and L. Kaatz, 2015: Embracing uncertainty: A case study examination of how climate change is shifting water utility planning. Water Utility Climate Alliance, American Water Works Association, Water Research Foundation, and Association of Metropolitan Water Agencies Final Tech. Rep., 92 pp., <https://www.wucaonline.org/assets/pdf/pubs-uncertainty.pdf>.
- Salas, J. D., 1993: Analysis and modeling of hydrologic time series. *Handbook of Hydrology*, D. R. Maidment, Ed., McGraw-Hill, 19.11–19.72.
- Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796–799, <https://doi.org/10.1126/science.1139540>.
- , and Coauthors, 2013: Real-time multi-model decadal climate predictions. *Climate Dyn.*, **41**, 2875–2888, <https://doi.org/10.1007/s00382-012-1600-0>.
- , and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate Atmos. Sci.*, **2**, 13, <https://doi.org/10.1038/s41612-019-0071-y>.
- Taylor, A. L., S. Dessai, and W. B. de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc.*, **373A**, 20140454, <https://doi.org/10.1098/rsta.2014.0454>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, [https://doi.org/10.1016/S0022-1694\(96\)03128-9](https://doi.org/10.1016/S0022-1694(96)03128-9).
- , M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook, 2016: Daymet: Daily surface weather data on a 1-km grid for North America, version 3. ORNL DAAC, accessed 9 May 2018, <https://doi.org/10.3334/ORNLDAAC/1328>.
- Towler, E., D. PaiMazumder, and J. Done, 2018: Toward the application of decadal climate predictions. *J. Appl. Meteor. Climatol.*, **57**, 555–568, <https://doi.org/10.1175/JAMC-D-17-0113.1>.
- , B. Rajagopalan, R. S. Summers, and D. Yates, 2010: An approach for probabilistic forecasting of seasonal turbidity threshold exceedance. *Water Resour. Res.*, **46**, W06511, <https://doi.org/10.1029/2009WR007834>.
- Udall, B., and J. Overpeck, 2017: The twenty-first century Colorado River hot drought and implications for the future. *Water Resour. Res.*, **53**, 2404–2418, <https://doi.org/10.1002/2016WR019638>.
- Vera, C., and Coauthors, 2010: Needs assessment for climate information on decadal timescales and longer. *Procedia Environ. Sci.*, **1**, 275–286, <https://doi.org/10.1016/j.proenv.2010.09.017>.
- WMO, 2020: WMO Lead Centre for Annual-to-Decadal Climate Prediction. Accessed 19 October 2020, <https://hadleyserver.metoffice.gov.uk/wmolc/>.
- Wood, A. W., T. Hopson, A. Newman, L. Brekke, J. Arnold, and M. Clark, 2016: Quantifying streamflow forecast skill elasticity to initial condition and climate prediction skill. *J. Hydrometeorol.*, **17**, 651–668, <https://doi.org/10.1175/JHM-D-14-0213.1>.
- , L. Woelders, and J. Lukas, 2020: Streamflow forecasting. *Colorado River Basin Climate and Hydrology: State of the Science*, J. Lukas and E. Payton, Eds., Western Water Assessment, 287–333.
- Yates, D., J. Sieber, D. Purkey, and A. Huber-Lee, 2005: WEAP21—A demand-, priority-, and preference-driven water planning model Part 1: Model characteristics. *Water Int.*, **30**, 487–500, <https://doi.org/10.1080/02508060508691893>.
- , and Coauthors, 2009: Climate driven water resources model of the Sacramento basin, California. *J. Water Resour. Plann. Manage.*, **135**, 303–313, [https://doi.org/10.1061/\(ASCE\)0733-9496\(2009\)135:5\(303\)](https://doi.org/10.1061/(ASCE)0733-9496(2009)135:5(303)).
- , K. Miller, R. L. Wilby, and L. Kaatz, 2015: Decision-centric adaptation appraisal for water management across Colorado's Continental Divide. *Climate Risk Manage.*, **10**, 35–50, <https://doi.org/10.1016/j.crm.2015.06.001>.
- Yeager, S. G., and Coauthors, 2018: Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **99**, 1867–1886, <https://doi.org/10.1175/BAMS-D-17-0098.1>.