Mitigating the Impacts of Climate Nonstationarity on Seasonal Streamflow Predictability in the U.S. Southwest

Flavio Lehner[1] ⓘD, Andrew W. Wood[1] ⓘD, Dagmar Llewellyn[2], Douglas B. Blatchford[3] ⓘD, Angus G. Goodbody[4] ⓘD, and Florian Pappenberger[5] ⓘD

[1]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [2]Bureau of Reclamation, Albuquerque Area Office, Albuquerque, NM, USA, [3]Bureau of Reclamation, Lower Colorado Regional Office, Boulder City, NV, USA, [4]National Water and Climate Center, Natural Resources Conservation Service, Portland, OR, USA, [5]Forecast Department, European Centre for Medium-Range Weather Forecasts, Reading, UK

Abstract Seasonal streamflow predictions provide a critical management tool for water managers in the American Southwest. In recent decades, persistent prediction errors for spring and summer runoff volumes have been observed in a number of watersheds in the American Southwest. While mostly driven by decadal precipitation trends, these errors also relate to the influence of increasing temperature on streamflow in these basins. Here we show that incorporating seasonal temperature forecasts from operational global climate prediction models into streamflow forecasting models adds prediction skill for watersheds in the headwaters of the Colorado and Rio Grande River basins. Current dynamical seasonal temperature forecasts now show sufficient skill to reduce streamflow forecast errors in snowmelt-driven regions. Such predictions can increase the resilience of streamflow forecasting and water management systems in the face of continuing warming as well as decadal-scale temperature variability and thus help to mitigate the impacts of climate nonstationarity on streamflow predictability.

1. Introduction

With growing populations and rising temperatures, the pressure on water resources in the southwestern United States (U.S.) is increasing and expected to continue to do so over the next decades (Reclamation, 2016). Water resources in California, Nevada, Arizona, Utah, Colorado, New Mexico, and Texas are currently almost entirely allocated for agricultural, industrial, and municipal uses and are heavily managed, with seasonal streamflow forecasts being a key tool used to inform this management. Seasonal streamflow forecasts for a range of lead times are among the most economically valuable streamflow predictions made in the United States and around the world, given their significance for water management (Hamlet et al., 2002; Raff et al., 2013).

Seasonal streamflow forecasts in the Upper Rio Grande river basin, for example, are used to predict the annual water delivery requirements between Colorado, New Mexico, and Texas under an interstate river allocation agreement, the Rio Grande Compact, to plan for water storage and to inform associated reservoir management decisions. The forecasts in combination with those decisions enable projections of the water supplies that will be available to farmers, which in turn can influence cropping decisions. In addition, supplemental water supply to the Upper Rio Grande basin is imported each year from the Colorado River system through transbasin diversions. Forecasts of the water available for diversion are used to estimate the portion of the imported water that will need to be purchased by the Federal government to support the needs of endangered species, as well as for planning of drinking water operations in major municipalities. On the much larger Colorado River system, as well, water supply forecasts issued in spring are essential to make reservoir storage and release decisions that help avoid shortage conditions in Lake Mead and Lake Powell, and that determine water and hydropower allocations affecting seven southwestern U.S. states. These decisions influence water and energy costs for major American cities such as Los Angeles, Las Vegas, and Phoenix, and major irrigation regions such as California's Imperial Valley and Arizona's Welton Mohawk Irrigation and Drainage District.

Although it is difficult to quantify the value of seasonal forecasts or the marginal value of forecast improvements, the value of the water managed using such forecasts rises well into the billions of dollars each year

(Hamlet et al., 2002; Pierce, 2010). In comparison, the costs of enhancements to operational water supply forecasting are small, especially when they represent an extension of the current approaches, similar to the cost-benefit ratio of improved flood forecasting (Pappenberger et al., 2015). In recent decades the western United States has seen strong hydroclimatic trends and decadal variability, leading to variable streamflow forecasting skill and a likelihood of suboptimal management decisions (Pagano & Garen, 2005). To better grapple with water resource management challenges arising from hydroclimate nonstationarity and increasing water demands, improved efficiency in water management practices is critically needed (Lins & Cohn, 2011; Milly et al., 2008; Steinschneider & Brown, 2012).

Operational seasonal streamflow forecasts in snowmelt-driven basins commonly derive skill from the stability of relationships between winter precipitation and snow water equivalent (SWE) with spring to summer melt season runoff (e.g., April–July streamflow). In some cases, but less commonly, additional predictability is found in observations of prior streamflow, soil moisture, and in climate indices such as El Niño–Southern Oscillation (Bell et al., 2017; Harpold et al., 2017; Kalra et al., 2013; Koster et al., 2010; Shukla & Lettenmaier, 2011; Wood et al., 2005). The simplest operational form of seasonal streamflow prediction relies on statistical models that quantify these relationships, such as principal component regression (PCR) models trained on observed in situ data records of ~30 years (Garen, 1992). These "water supply forecasts" (WSFs) have traditionally been made beginning in January of the same year with updates on the first day of each month to incorporate new precipitation and SWE observations (Pagano, Wood, Ramos, et al., 2014). Operational forecasts are published by regional River Forecasting Centers and the U.S. Department of Agriculture National Resources Conservation Service (NRCS). A second common form of seasonal streamflow prediction involves the use of dynamic watershed models to predict future watershed states and fluxes (Day, 1985; Pagano, Wood, Werner, et al., 2014).

The skill of statistical WSFs varies with lead time and also on decadal time scales, with basins such as the Upper Colorado River (UC) and Upper Rio Grande (URG) showing declining skill since the 1980s (Pagano et al., 2004). While extensive research has been conducted on how to improve seasonal streamflow forecasting systems (Crochemore et al., 2016; Mendoza et al., 2017; Moradkhani et al., 2004; Wood & Lettenmaier, 2006, 2008), the reasons for decadal variations in skill of a fixed forecasting system remain relatively elusive. Pagano and Garen (2005) argue that these skill variations originate primarily from interannual to decadal climate variations, rather than basin-specific processes or human interference. As such, successful prediction of interannual to decadal climate variability has the potential to stabilize streamflow forecasting skill.

Besides decadal climate variability, southwestern U.S. water resources are also sensitive to the influence of anthropogenically forced climate change, be it via temperature, precipitation, or atmospheric circulation changes (Barnett et al., 2005; Christensen et al., 2004; Lettenmaier & Gan, 1990; Mote et al., 2005). For semiarid and snowmelt driven basins such as the UC and URG, numerous studies have indicated that increasing temperature decreases streamflow (Christensen et al., 2004; Griffin & Friedman, 2017; Lehner et al., 2017; Nowak et al., 2012; Udall & Overpeck, 2017; Vano et al., 2012; Woodhouse et al., 2016). Specifically, runoff efficiency—a metric indicating the fraction of precipitation that ends up as streamflow—is more likely to be low when temperatures are above average (Lehner et al., 2017; Nowak et al., 2012). As a consequence, the relationship between winter moisture accumulation (precipitation and SWE) and summer streamflow is evidently nonstationary and can be influenced by temperature.

The influence of temperature on runoff efficiency is problematic for WSFs in light of their underlying stationarity assumptions with regard to the background climate during the forecast period. Statistical models using observed accumulated precipitation and SWE at the start of the forecast without additional temperature information for the forecast period would underpredict streamflow for cool forecast periods and overpredict streamflow for warm forecast periods, in part because they do not include the information of the secular warming trend and associated evaporation losses over the entire period.

Here we investigate (1) recent hydroclimate trends and streamflow forecast errors in the study region, the URG and parts of the UC, (2) the seasonal predictability of temperature over this region, and (3) whether including predicted temperatures in WSFs improves seasonal streamflow forecasting skill. To that end, we generate WSFs via the current operational strategy, termed "baseline forecast," as well as WSFs that include seasonal temperature forecasts as a predictor, termed "temperature-aided forecast". The comparison of the two approaches enables us to assess the potential to improve streamflow forecasting skill by including

temperature forecasts, as well as the sufficiency of current operational temperature forecasts for this purpose. Section 2 introduces the data and methods used, section 3 presents the results, and section 4 discusses their wider implications.

2. Data and Methods

2.1. Streamflow, Precipitation, Snow Water Equivalent, and Temperature Data Sets

Estimates of naturalized monthly streamflow at a number of gages across the UC and URG are obtained from the NRCS; the gages are marked with circles in Figure 1a and are listed in Table S1 in the supporting information. For each gage and year from 1987 to 2016, the total streamflow for the respective forecasting "target period" (e.g., April–July cumulative flow) is calculated. Observations of water year-to-date cumulative precipitation and instantaneous SWE at 1 January, 1 February, 1 March, 1 April, and 1 May are extracted from the same snow telemetry monitoring (SNOTEL) stations as used in the operational forecasting by NRCS, but only if they cover the entire hindcasting period 1987–2016 (triangles in Figure 1a; see also Table S1); this is to ensure consistency and reproducibility across the hindcasting period. The year 1987 is chosen as a start year because it offers continuous streamflow and SNOTEL measurements across all gages considered here. Monthly mean temperature is taken from the Parameter Elevation Regression on Independent Slopes Model (PRISM) data set (Daly et al., 2008) averaged over the box indicated in Figure 1 (35.5–39.5°N, 108.5–105.0°W). Precipitation used to calculate runoff efficiency in Figure 1b is taken from PRISM as well, summed up over the watersheds upstream of Rio Grande at Otowi Bridge, San Juan at Bluff, and Gunnison at Grand Junction.

2.2. Seasonal Temperature Forecasts

Seasonal temperature forecasts are derived from eight initialized coupled climate models that produce seasonal climate forecasts (Table S2): the North American Multimodel Ensemble (NMME; Kirtman et al., 2014), which comprises seven models, and the System 4 seasonal forecasting model from the European Center for Medium-Range Weather Forecast (ECMWF; Molteni et al., 2011). In their current configuration, these models issue forecasts each month for lead times of up to 12 months with various numbers of ensemble members (10–51). Since we are interested in extracting the seasonally predictable signal, we use each model's ensemble mean (rather than all its individual ensemble members) of monthly mean 2 m temperature hindcasts issued from January 1987 to May 2016, averaged over the area indicated in Figure 1a. We then use an equal-weights multimodel mean across the eight models, since we found this method to perform, in terms of correlation with observed temperature, as well as or better than other weighting schemes in cross-validation across issue dates and lead times of interest (we tested a performance-weighted multimodel mean and an equal-weights mean of the overall three best models CFSv2, NASA, and ECMWF; not shown). For each streamflow forecast issue month (1 January, 1 February, etc.), temperature is averaged from that issue month until the end of the main runoff period (July). Alternatives to this choice were tested, such as using spring (March–May) average temperature or the average over the next or the next 2 months after issue date, but were found to be inferior (not shown).

2.3. Streamflow Forecasting Procedure

The marginal benefit of including seasonal temperature information in WSFs can be evaluated through benchmarking the performance of enhanced WSF models against models based on the current operational forecast practice. We mimic the operational forecasting procedure of the NRCS's operational WSF by using SNOTEL data in a principal component regression (PCR) trained on 30 years (1987–2016) of observed naturalized streamflow of the respective target period (Garen, 1992), hereafter "baseline forecast". Before use in the PCR, all predictors are standardized (subtraction of mean and division by standard deviation) and streamflow is seminormalized via a square root transformation, as is consistent with NRCS practice. The number of principal components (PCs) retained is determined through an iterative process as described in Garen (1992). Specifically, individual PCs are used in a linear regression and the significance of the regression coefficients is determined via a t test; only PCs are retained that result in significant regression coefficients and that show a physically plausible relationship with streamflow (i.e., positive coefficients, indicating that high precipitation and SWE typically leads to high streamflow and vice versa). In our case, one PC is retained for all streamflow gages, consistent with Harpold et al. (2017), who also duplicated the NRCS's WSF. For each forecast issue date, forecasts are cross-validated by training the model on 29 of the 30 years and forecast the
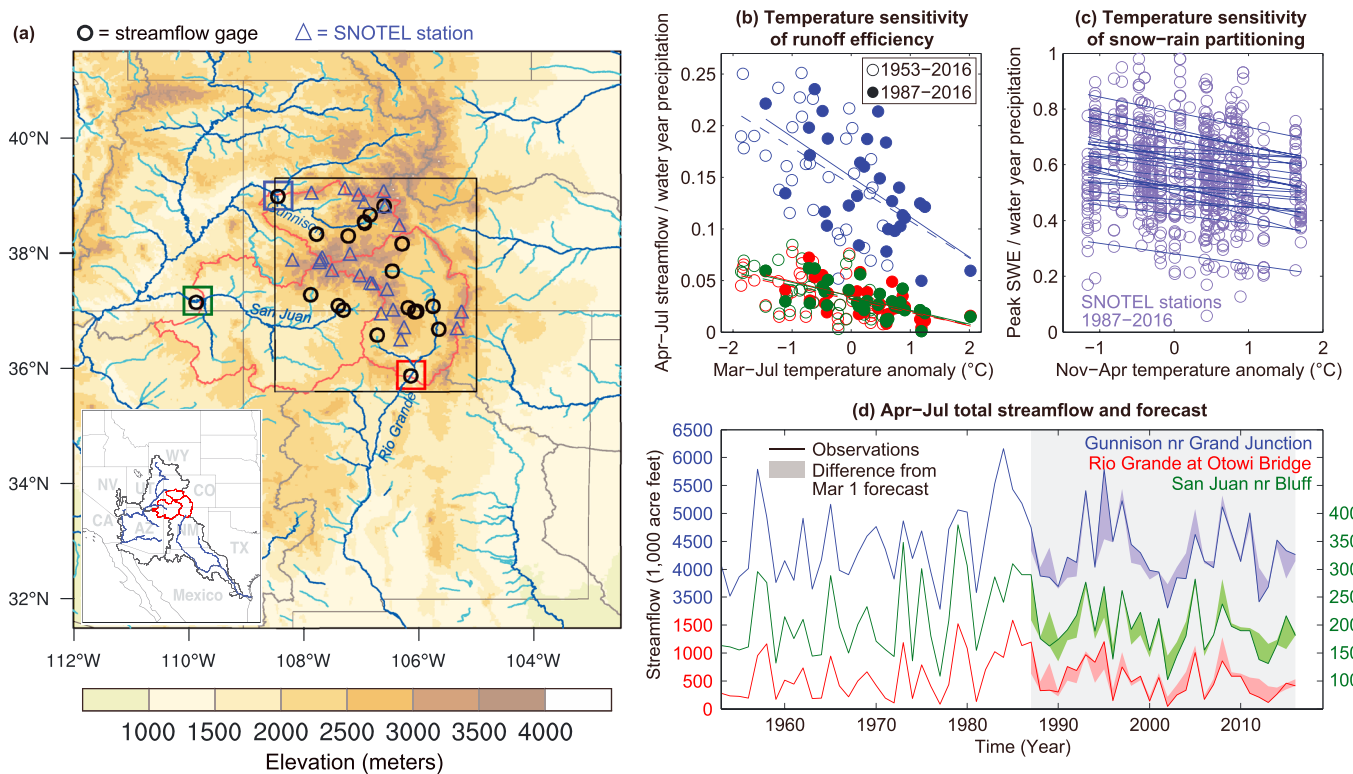
Figure 1. (a) Map showing the main rivers, basins, (circles) streamflow gages, and (triangles) SNOTEL stations analyzed in this study. (b) Runoff efficiency—spring-summer streamflow divided by water year precipitation—as a function of spring-summer temperature for three selected gages marked with colored boxes in Figure 1a. (c) Snow-rain partitioning—peak snow water equivalent (SWE) divided by water year precipitation—for all SNOTEL stations analyzed in this study (each linear trend line is for one SNOTEL station) as a function of winter-spring temperature. (d) Observed and forecasted streamflow for the three selected gages; solid lines are the observed streamflow, while colored shading indicates the difference between the observed and forecasted streamflow, that is, the larger the shading the larger the forecast error; gray shading indicates time period analyzed in this study. See text for more details on data sets.

remaining (out-of-sample) year, loop through all 30 years to evaluate performance. Note that our baseline forecast likely differs slightly from the officially published NRCS forecast over the past decades, since those may also include additional but noncontinuous snow course information and/or newer SNOTEL data. As discussed above, for consistency across watersheds, we only use data sets of consistent record length (1987–2016).

We then reforecast the same time period using the same information but add the ensemble mean temperature anomaly of the eight seasonal forecasting models as an additional predictor to the PCR (hereafter "temperature-aided forecast"). For a given year and forecast issue date (e.g., 1 January, 1987), the mean temperature prediction from the forecast issue date to the end of July is averaged over the box indicated in Figure 1a. For all gages, the regression coefficients derived from the PCR are such that precipitation and SWE always exhibit a positive relationship with streamflow, and temperature always a negative one, indicating a physically plausible interaction of precipitation, SWE, and temperature in describing streamflow. The same rules for PC retention are applied and one PC was retained in all cases.

2.4. Skill Metrics

Prediction skill for the baseline and temperature-aided streamflow forecast is calculated via a leave-one-out cross validation from 1987 to 2016. Each year between 1987 and 2016 is hindcasted with a principal component regression model that has been calibrated on the remaining 29 years of data, and the resulting time series of 30 streamflow predictions are verified against the corresponding observations.

We quantify forecast skill using the following metrics: (i) correlation, (ii) relative root-mean-square error (rRMSE, in %), (iii) the Brier Skill Score (BSS) for streamflow <33rd percentile, and (iv) Continuous Ranked Probability Skill Score (CRPSS; Hersbach, 2000). Correlation and rRMSE describe how well the model predicts the variability and the absolute values, respectively, of the observed time series. The third metric provides insight into the ability of the model to predict dry conditions relevant to droughts in the U.S. Southwest,
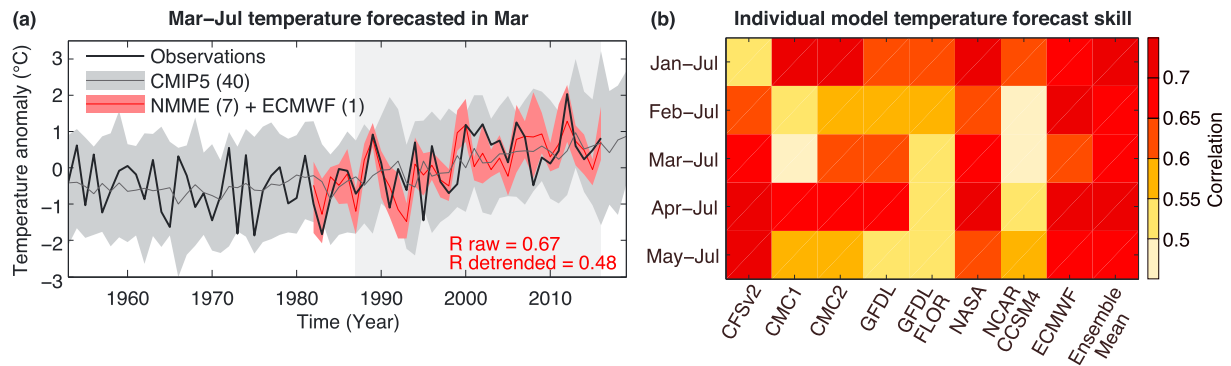
Figure 2. (a) March–July mean temperature anomalies relative to 1982–2016 from observations, 40 CMIP5 models, and seasonal prediction models (NMME + ECMWF), averaged over the box indicated in Figure 1a. The red line is the mean across NMME-ECMWF models, the gray line is the mean across CMIP5 models, and the black line is observations. Shading indicates the 5–95% range. (b) Correlation between observed and forecasted temperature for different temperature targets and seasonal prediction models for 1982–2016. Forecasts are initialized at the start of each predicted period. All correlations are significant at 95% confidence.

and the fourth metric, which measures the ability of the forecast model to correctly predict the cumulative distribution function of the observed streamflow data, is used to quantify probabilistic prediction skill.

Since the skill metrics BSS and CRPSS rely on probabilistic forecasts, we derive exceedance probabilities from the standard error of the forecasts, consistent with NRCS' approach (Garen, 1992). Both BSS and CRPSS are typically expressed as skill relative to a certain reference forecast (typically persistence or climatology). Here we express them relative to the baseline forecast to emphasize the improvement relative to the current operational approach.

3. Results

3.1. Hydroclimate Trends and Streamflow Forecast Errors

Recent hydroclimate trends in the UC and URG headwaters are illustrated by plotting the runoff efficiency as a function of temperature anomalies for streamflow gages at the outflow of the headwaters of the Gunnison, San Juan, and Rio Grande (Figure 1b; these three gages are representative of the dynamics at other gages, see Figure S1). A clear temperature sensitivity exists, leading to relatively reduced streamflow under positive temperature anomalies. Even in the absence of a strong precipitation trend, higher temperatures are shifting the partitioning of precipitation from snow to rain, a phenomenon that is detectable at virtually all SNOTEL stations in the region (Figure 1c), thereby changing the peaks and timing of both snowmelt and runoff. Higher temperatures also allow for more evaporative loss between when the snow falls and when the water arrives at the streamflow gages downstream, which is a key hydrologic dynamic leading to forecast errors.

Relatively persistent forecast errors are confirmed by the forecast record in the UC and URG: streamflow gage records in these two basins show a tendency to be underpredicted in the 1980s and 1990s and overpredicted in the 2000s and 2010s (Figures 1d and S1). While these forecast errors are in part related to unusually wet springs and summers in the 1980–1990s and unusually dry springs and summers in the 2000–2010s, there exists evidence that streamflow in recent years was lower than expected from precipitation deficits alone (Lehner et al., 2017; Woodhouse et al., 2016), pointing to a simultaneous influence of temperature on streamflow and thus on forecast error. This theory is further corroborated by a significant correlation of streamflow forecast error with both anomalous precipitation and temperature after the forecast issue date (Figure S2). This relationship holds even when the natural correlation between precipitation and temperature is accounted for, a result consistent with earlier studies (Harding et al., 2012).

3.2. Temperature Forecast Skill

While uncertainty in multidecadal projections of precipitation in the U.S. Southwest remains high, climate models such as those included in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) project future temperature increases (Figure 2a) with far more certainty (van Oldenborgh et al., 2013). Similarly, dynamical seasonal climate prediction models, such as the eight models from the NMME and

ECMWF are more skillful in predicting temperature than precipitation (Becker et al., 2014; Slater et al., 2016). The ensemble mean across these eight seasonal forecasting models captures the observed warming trend of recent decades as well as part of the interannual variability of spring-to-summer temperature over the UC and URG headwaters region at lead times of up to 5 months, showing significant correlations ranging between 0.65 and 0.75 (Figures 2a and 2b). The combination of these two results leads to a usable temperature forecast skill in the context of streamflow prediction in this region. The ECMWF model is the best performing individual model overall, although not necessarily for every lead times and not necessarily when compared to the multi-model mean across all eight models.

3.3. Improved Streamflow Forecast Skill

We find that augmenting the baseline forecasting approach through the use of temperature predictors adds prediction skill across the majority of streamflow gages and issue dates in the study region, which is representative of snowmelt-influenced watersheds in many parts of the western United States. These benefits are illustrated through the skill difference between the baseline and temperature-aided forecasts for all skill metrics considered (Figure 3). The median relative improvement across gages and skill metrics is between 1% and 5% with some spread across gages. The vast majority of these improvements are statistically significant in light of sampling uncertainty (see section 3.4). However, the probabilistic skill for drought conditions (BSS) is improved less consistently than the other skill metrics. All four skill metrics indicate larger improvements for later issue dates, which likely results from a combination of better temperature forecast skill at shorter lead times and the potential for stronger temperature anomaly signals due to a shorter averaging period (e.g., May–July versus January–July).

When considering the median skill across gages within each basin, improvements tend to be larger in the Rio Grande and San Juan than in the Gunnison. The variations of forecast improvements across gages reflects the different temperature sensitivity of catchment hydrology in different locations. The sensitivity of spring run-off to temperature is affected by factors such as the basin distribution of elevation and aspect, vegetation, and land cover (Male & Gray, 1981), making it difficult to disentangle the reasons for an individual forecast's improvement using a statistical model only. No relationship between magnitude of skill improvement and basin elevation is found (not shown).

We also calculate the theoretical skill improvement resulting from using the actually observed temperature and found it overall to be only marginally higher than with the temperature-aided forecast based on predicted temperature (Figures 3b and 3c). This indicates that the majority of the temperature information that adds skill to WSF can indeed be extracted from seasonal prediction models. Since temperature in this region over the period 1987–2016 shows a strong positive trend, the question arises, How much of the added skill is attributable to the trend alone? Using the observed linear temperature trend from 1987 to 2016 as a predictor in the WSF model (thereby excluding any interannual variability that might be predictable by seasonal prediction models), we show that the trend alone adds most of the skill that originates from the seasonally forecasted temperatures (Figure 3a and 3b).

Finally, we repeat the forecasting using the temperature forecasts from the ECMWF model only, since it is the best performing individual model (Figure 2b), and from the seven NMME models only (i.e., without ECMWF). Interestingly, we found the streamflow forecasting skill to be roughly equal in all three cases (Figure S3). This suggests that temperature forecasts from ECMWF model contain about as much information, with regard to streamflow forecasting, as the seven NMME models combined.

3.4. Robustness of Forecast Skill Improvements

The skill is improved for the majority of the total of 100 possible forecasts (20 gages × 5 issue dates). For correlation, rRMSE, and CRPSS, 99% of forecasts are improved, with 98-100% of those significantly. For BSS, only 62% of all forecasts are improved, 95% of those significantly (see also Table S3). Significance is established through a Monte Carlo approach in which all forecasts and the associated skill score calculations are repeated 1,000 times on 30 year samples constructed from bootstrapping the original 30 years with replacement. If the 95th percentile of this distribution of skill scores shows an improvement, the skill improvement is considered significant at the 95% confidence level.
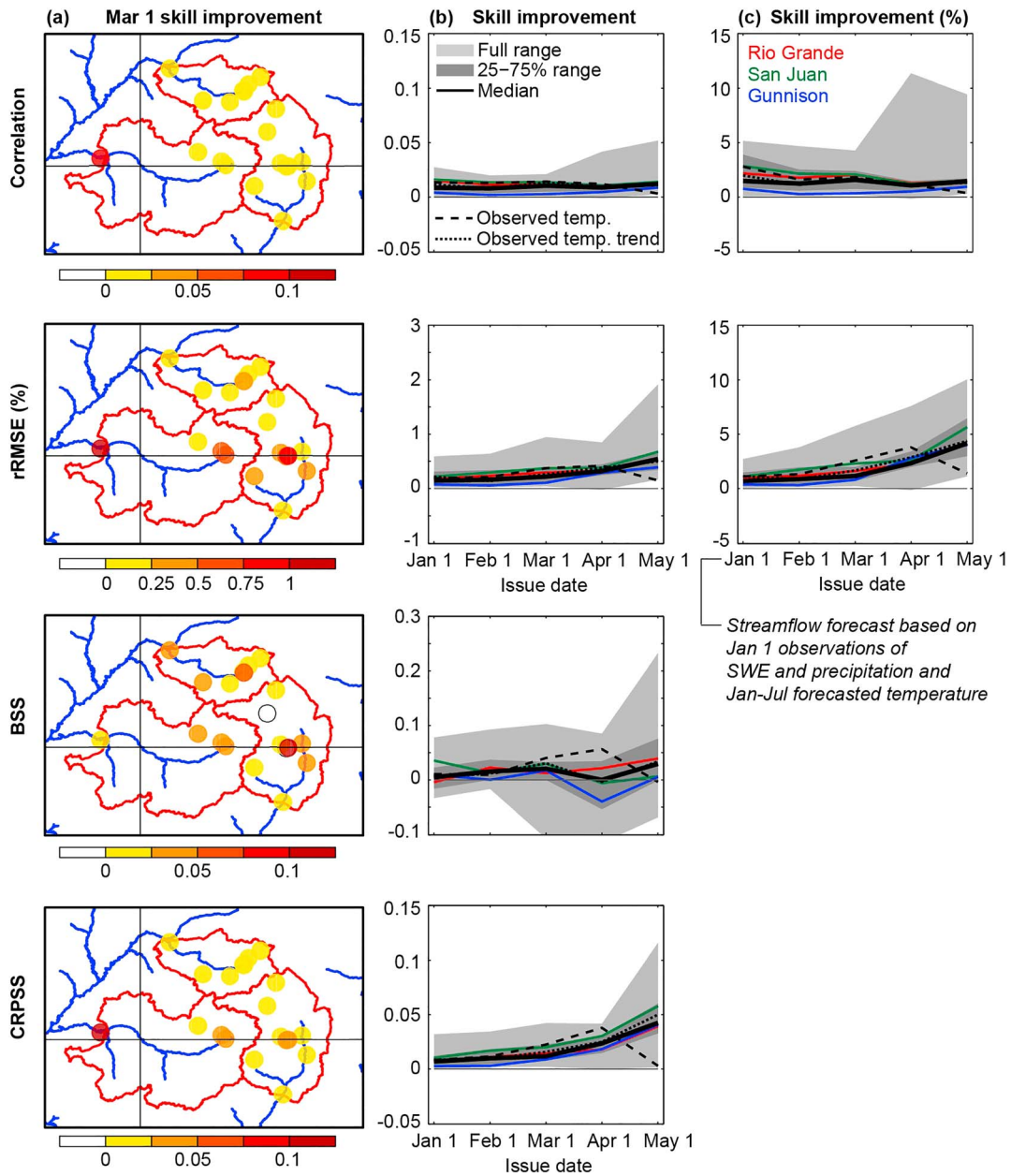
Figure 3. (a) Absolute skill improvement of the temperature-aided forecast relative to the baseline forecast at individual gages for issue date 1 March as an illustrative example. (b) Absolute skill improvement for all gages as a function of issue date. (c) Relative skill improvement for all gages as a function of issue date. Solid lines are the median across (black) all gages and (colors) the three basins. Dashed line is the median across all gages when using observed temperature, mimicking the hypothetical case where the future temperature is known at the time of forecast issue, and dotted line is the median when using only the linear trend of observed temperature.

4. Discussion and Conclusions

The skill improvement demonstrated here for seasonal streamflow forecasts in the Upper Rio Grande and Upper Colorado River basins can be of significant value to State and Federal water managers, which, in turn, can benefit water users throughout these basins (C. Donnelly and C. Cotton, personal communication, 2017). Despite its limited spatial extent, the study here is of relevance for other snowmelt-driven basins across the United States and the world, since streamflow forecast skill in such basins is often driven by the same temperature-sensitive processes.

We show that current seasonal climate prediction models are skillful in forecasting both the long-term trends and interannual variability of seasonal temperatures for this region. This temperature information adds skill to existing "water supply forecasts" (WSFs), mitigating some of the forecast errors introduced through climate nonstationary, and moving the WSFs closer to their maximally expected forecast skill based on relationships between observed snow, precipitation, and temperature. Additional predictability might be available once seasonal precipitation forecasts become more skillful.

For the statistical WSFs shown here, the proposed extension involves accessing and incorporating temperature predictions into existing statistical forecasting models. Conventional forecasting approaches based on hydrologic models (such as Ensemble Streamflow Prediction, or ESP, a popular operational method that is not discussed in this paper) are also commonly dependent on climate stationarity assumptions and thus are also likely to benefit from additional temperature forecast information. Fortunately, many techniques for inclusion of conditional climate information have been described in the literature over the last several decades for both statistical and model-based forecasting (e.g., Beckers et al., 2016; Mendoza et al., 2017; Werner et al., 2004; see also the special issue of Wetterhall et al., 2017), including examples of using NMME and ECMWF (Crochemore et al., 2016; Mo & Lettenmaier, 2014; Thober et al., 2015; Yuan et al., 2013). It may be impossible to protect or increase streamflow prediction skill in all locations in the face of a nonstationary climate, but expanding the use of model-based seasonal climate predictions, and particularly temperature forecasts, appears to be one pragmatic strategy for hydroclimates that are similar to the U.S. Southwest.

Despite the evidence of forecast skill improvement through inclusion of temperature, this study does not support detailed conclusions regarding the hydrologic processes that underpin changes in prediction skill, as the temperature influence on streamflow can be dampened or amplified due to other effects and non-linear interactions (e.g., related to groundwater use or vegetation alterations). Our focus on minimally impaired gages in headwater locations aims to circumvent this issue, but we cannot exclude all possibilities of processes amplifying or canceling each other. Similarly, using low-dimensional statistical models only, we are unable to disentangle why certain gages show greater improvement than others. Process-based observation and modeling studies tackling this question may therefore be a valuable next step for the hydrologic forecasting community.

References

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303–309. https://doi.org/10.1038/nature04141

Becker, E., Van den Dool, H., & Zhang, Q. (2014). Predictability and forecast skill in NMME. *Journal of Climate*, *27*(15), 5891–5906. https://doi.org/10.1175/JCLI-D-13-00597.1

Beckers, J. V. L., Weerts, A. H., Tijdeman, E., & Welles, E. (2016). ENSO-conditioned weather resampling method for seasonal ensemble streamflow prediction. *Hydrology and Earth System Sciences*, *20*(8), 3277–3287. https://doi.org/10.5194/hess-20-3277-2016

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., & Scaife, A. A. (2017). A national-scale seasonal hydrological forecast system: Development and evaluation over Britain. *Hydrology and Earth System Sciences Discussions*, 1–19. https://doi.org/10.5194/hess-2017-154

Bureau of Reclamation (2016). Climate Change Adaptation Strategy. (33 pp.). Retrieve from https://www.usbr.gov/climate/docs/2016ClimateStrategy.pdf

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., & Palmer, R. N. (2004). The effects of climate change on the hydrology and water resources of the Colorado River basin. *Climatic Change*, *62*(1-3), 337–363. https://doi.org/10.1023/B:CLIM.0000013684.13621.1f

Crochemore, L., Ramos, M. H., & Pappenberger, F. (2016). Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *20*(9), 3601–3618. https://doi.org/10.5194/hess-20-3601-2016

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., … Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, *111*(2), 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

Garen, D. C. (1992). Improved techniques in regression-based streamflow volume forecasting. *Journal of Water Resources Planning and Management*, *118*(6), 654–670. https://doi.org/10.1061/(ASCE)0733-9496(1992)118:6(654)

Griffin, E. R., & Friedman, J. M. (2017). Decreased runoff response to precipitation, little Missouri River basin, northern Great Plains, USA. *Journal of the American Water Resources Association*, *53*(3), 576–592. https://doi.org/10.1111/1752-1688.12517

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management*, *128*(2), 91–101. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91)

Harding, B. L., Wood, A. W., & Prairie, J. R. (2012). The implications of climate change scenario selection for future streamflow projection in the upper Colorado River basin. *Hydrology and Earth System Sciences*, *16*(11), 3989–4007. https://doi.org/10.5194/hess-16-3989-2012

Harpold, A. A., Sutcliffe, K., Clayton, J., Goodbody, A., & Vazquez, S. (2017). Does including soil moisture observations improve operational streamflow forecasts in snow-dominated watersheds? *Journal of the American Water Resources Association*, *53*(1), 179–196. https://doi.org/10.1111/1752-1688.12490

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, *15*(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015%3C0559:DOTCRP%3E2.0.CO;2

Kalra, A., Miller, W. P., Lamb, K. W., Ahmad, S., & Piechota, T. (2013). Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River basins. *Hydrological Processes*, *27*(11), 1543–1559. https://doi.org/10.1002/hyp.9236

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L. III, Paolino, D. A., Zhang, Q., … Wood, E. F. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Koster, R. D., Mahanama, S. P. P., Livneh, B., Lettenmaier, D. P., & Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience*, *3*(9), 613–616. https://doi.org/10.1038/ngeo944

Lehner, F., Wahl, E. R., Wood, A. W., Blatchford, D., & Llewellyn, D. (2017). Assessing recent declines in Upper Rio Grande River runoff efficiency from a paleoclimate perspective. *Geophysical Research Letters*, *44*, 4124–4133. https://doi.org/10.1002/2017GL073253

Lettenmaier, D. P., & Gan, T. Y. (1990). Hydrologic sensitivities of the Sacramento-San Joaquin River Basin, California, to global warming. *Water Resources Research*, *26*(1), 69–86. https://doi.org/10.1029/WR026i001p00069

Lins, H. F., & Cohn, T. A. (2011). Stationarity: Wanted dead or alive? *Journal of the American Water Resources Association*, *47*(3), 475–480. https://doi.org/10.1111/j.1752-1688.2011.00542.x

Male, D. H., & Gray, D. M. (Eds.), (1981). *Handbook of snow: Principles, processes, management and use*, (p. 776). Caldwell, New Jersey: The Blackburn Press.

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., … Arnold, J. R. (2017). An intercomparison of approaches for improving predictability in operational seasonal streamflow forecasting. *Hydrology and Earth System Sciences Discussions*, 1–37. https://doi.org/10.5194/hess-2017-60

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

Mo, K. C., & Lettenmaier, D. P. (2014). Hydrologic prediction over the conterminous United States using the National Multi-Model Ensemble. *Journal of Hydrometeorology*, *15*(4), 1457–1472. https://doi.org/10.1175/JHM-D-13-0197.1

Molteni, F., Stockdale, T., Balmaseda, M. A., Balsamo, G., Buizza, R., Ferranti, L., … Vitart, F. (2011). The new ECMWF seasonal forecast system (System 4). Tech. Memo. ECMWF, 49.

Moradkhani, H., Hsu, K., Gupta, H. V., & Sorooshian, S. (2004). Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *Journal of Hydrology*, *295*(1-4), 246–262. https://doi.org/10.1016/j.jhydrol.2004.03.027

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of the American Meteorological Society*, *86*(1), 39–49. https://doi.org/10.1175/BAMS-86-1-39

Nowak, K., Hoerling, M., Rajagopalan, B., & Zagona, E. (2012). Colorado River Basin hydroclimatic variability. *Journal of Climate*, *25*(12), 4389–4403. https://doi.org/10.1175/JCLI-D-11-00406.1

Pagano, T. C., & Garen, D. (2005). The recent increase in Western US streamflow variability and persistence. *Journal of Hydrometeorology*, *6*(2), 173–179. https://doi.org/10.1175/JHM410.1

Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of official western U.S. seasonal water supply outlooks, 1922–2002. *Journal of Hydrometeorology*, *5*(5), 896–909. https://doi.org/10.1175/1525-7541(2004)005%3C0896:EOOWUS%3E2.0.CO;2

Pagano, T. C., Wood, A. W., Ramos, M. H., Cloke, H. L., Pappenberger, F., Clark, M. P., … Verkade, J. S. (2014). Challenges of operational river forecasting. *Journal of Hydrometeorology*, *15*(4), 1692–1707. https://doi.org/10.1175/JHM-D-13-0188.1

Pagano, T., Wood, A., Werner, K., & Tama-Sweet, R. (2014). Western U.S. water supply forecasting: A tradition evolves. *Eos, Transactions American Geophysical Union*, *95*, 28–29. https://doi.org/10.1002/2014EO030007

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Environmental Science and Policy*, *51*, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Pierce, J. A. S. (2010). *A measure of snow: Case studies of the snow survey and water supply forecasting program* (p. 109). Salt Lake City, UT: USDA. Retrieve from https://www.wcc.nrcs.usda.gov/ftpref/downloads/factpub/MeasureofSnowFullReport.pdf

Raff, D., Brekke, L., Werner, K., Wood, A., & White, K. (2013). Short-term water management decisions: User needs for improved climate, weather, and hydrologic information. (231 pp.). Retrieve from http://www.ccawwg.us/

Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences*, *15*(11), 3529–3538. https://doi.org/10.5194/hess-15-3529-2011

Slater, L. J., Villarini, G., & Bradley, A. A. (2016). Evaluation of the skill of North-American Multi-Model Ensemble (NMME) Global Climate Models in predicting average and extreme precipitation and temperature over the continental USA. *Climate Dynamics*. https://doi.org/10.1007/s00382-016-3286-1

Steinschneider, S., & Brown, C. (2012). Dynamic reservoir management with real-option risk hedging as a robust adaptation to nonstationary climate. *Water Resources Research*, *48*, W05524. https://doi.org/10.1029/2011WR011540

Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D., & Samaniego, L. (2015). Seasonal soil moisture drought prediction over Europe using the North American Multi-Model Ensemble (NMME). *Journal of Hydrometeorology*, *16*(6), 2329–2344. https://doi.org/10.1175/JHM-D-15-0053.1

Udall, B., & Overpeck, J. (2017). The 21st century Colorado River hot drought and implications for the future. *Water Resources Research*, *53*, 2404–2418. https://doi.org/10.1002/2016WR019638

van Oldenborgh, G., Collins, M., Arblaster, J., Christensen, J., Marotzke, J., Power, S., … Zhou, T. (2013). Atlas of global and regional climate projections. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1311–1394). Cambridge, UK, and NY: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.029

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

Werner, K., Brandon, D., Clark, M., & Gangopadhyay, S. (2004). Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. *Journal of Hydrometeorology*, *5*(6), 1076–1090. https://doi.org/10.1175/JHM-381.1

Wetterhall, F., Pechlivanidis, I. G., Ramos, M.-H., Wood, A. W., Wang, Q. J., Zehe, E., & Ehret, U. (2017). Special issue: Sub-seasonal to seasonal hydrological forecasting. *Hydrology and Earth System Sciences*, *21*(9), 4681–4691. https://doi.org/10.5194/hess-21-4681-2017

Wood, A. W., Kumar, A., & Lettenmaier, D. P. (2005). A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States. *Journal of Geophysical Research*, *110*, D04105. https://doi.org/10.1029/2004JD004508

Wood, A. W., & Lettenmaier, D. P. (2006). A test bed for new seasonal hydrologic forecasting approaches in the western United States. *Bulletin of the American Meteorological Society*, *87*(12), 1699–1712. https://doi.org/10.1175/BAMS-87-12-1699

Wood, A. W., & Lettenmaier, D. P. (2008). An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, *35*, L14401. https://doi.org/10.1029/2008GL034648

Woodhouse, C. A., Pederson, G. T., Morino, K., McAfee, S. A., & McCabe, G. J. (2016). Increasing influence of air temperature on upper Colorado River streamflow. *Geophysical Research Letters*, *43*, 2174–2181. https://doi.org/10.1002/2015GL067613

Yuan, X., Wood, E. F., Roundy, J. K., & Pan, M. (2013). CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States. *Journal of Climate*, *26*(13), 4828–4847. https://doi.org/10.1175/JCLI-D-12-00683.1

Erratum

In the originally published version of this article, there were errors in the text due to an error in the forecasting procedure. The third sentence in the first paragraph of Section 2.3 should have read: "Before use in the PCR, all predictors are standardized (subtraction of mean and division by standard deviation) and streamflow is seminormalized via a square root transformation, as is consistent with NRCS practice." The third sentence in the first paragraph of Section 3.3 should have read: "The median relative improvement across gages and skill metrics is between 1% and 5% with some spread across gages." The fifth sentence in the first paragraph of Section 3.3 should have read: "However, the probabilistic skill for drought conditions (BSS) is improved less consistently than the other skill metrics." The first sentence in the second paragraph of Section 3.3 should have read: "When considering the median skill across gages within each basin, improvements tend to be larger in the Rio Grande and San Juan than in the Gunnison." The fourth and fifth sentences of the 3rd paragraph of Section 3.3 should have read: "Using the observed linear temperature trend from 1987 to 2016 as a predictor in the WSF model (thereby excluding any interannual variability that might be predictable by seasonal prediction models), we show that the trend alone adds most of the skill that originates from the seasonally forecasted temperatures (Figure 3a and 3b)." The second sentence in the first paragraph of Section 3.4 should have read: "For correlation, rRMSE, and CRPSS, 99% of forecasts are improved, with 98-100% of those significantly. For BSS, only 62% of all forecasts are improved, 95% of those significantly (see also Table S3)." There were additional errors in Figure 3 and the Supporting information. These errors have since been corrected, and this version may be considered the authoritative version of record.