Article type : Technical Paper

Developing Sub-seasonal to Seasonal Climate Forecast Products for Hydrology and Water Management

Sarah A. Baker, Andrew W. Wood, and Balaji Rajagopalan

Department of Civil, Environmental, and Architectural Engineering (**Baker**, **Rajagopalan**), University of Colorado Boulder, Boulder, Colorado, USA; and Research Applications Laboratory (**Wood**), National Center for Atmospheric Research, Boulder, Colorado, USA (Correspondence to Baker: sarah.a.baker@colorado.edu).

Research Impact Statement: The post-processing of climate forecasts to watershed scales increases their relevance and potential utility for water management.

ABSTRACT: We describe a new effort to enhance climate forecast relevance and usability through the development of a system for evaluating and displaying real-time sub-seasonal to seasonal (S2S) climate forecasts on a watershed scale. Water managers may not use climate forecasts to their full potential due to perceived low skill, mismatched spatial and temporal resolutions, or lack of knowledge or tools to ingest data. Most forecasts are disseminated as large-domain maps or gridded datasets and may be systematically biased relative to watershed climatologies. Forecasts presented on a watershed scale allow water managers to view forecasts for their specific basins, thereby increasing the usability and relevance of climate forecasts. This paper describes the formulation of S2S climate forecast products based on the Climate Forecast System version 2 (CFSv2) and the North American Multi-model Ensemble (NMME). Forecast products include bi-weekly CFSv2 forecasts, and monthly and seasonal NMME forecasts. Precipitation and temperature forecasts are aggregated spatially to a USGS HUC-4 watershed scale. Forecast verification reveals appreciable skill in the first two bi-weekly periods (weeks 1-2 and 2-3) from CFSv2, and usable skill in NMME month 1 forecast with varying skills at longer lead times dependent on the season. Application of a bias-correction technique (quantile mapping) eliminates forecast bias in the CFSv2 reforecasts, without adding significantly to correlation skill.

(**KEYWORDS:** watershed management; geospatial analysis; sub-seasonal to seasonal; climate forecasts; CFSv2; NMME; precipitation; temperature.)

INTRODUCTION & BACKGROUND

Hydrologists and water managers make many operational decisions on a sub-seasonal to seasonal (S2S) time scale, but under-utilize climate prediction to inform decision making from a quantitative standpoint. Surveys indicate that water managers are reluctant to use climate forecast due to perceived poor reliability of forecasts, mismatched temporal or spatial scale, institutional reasons such as traditional reliance on built infrastructure, organization or regulatory restraints, and risk aversion (Callahan *et al.*, 1999; Kirchhoff *et al.*, 2013; Rayner *et al.*, 2005; White *et al.*, 2017). Water managers may be unaware of sources of seasonal climate forecasts or lack the skill set and resources to ingest forecasts in a usable format, especially managers at smaller utilities (Bolson *et al.*, 2013). Issues presented in these academic surveys can be addressed through a closer relationship between forecast producer and user, increased institutional flexibility, and demonstration of effective climate forecast skill and use (Dilling and Lemos, 2011; Feldman and Ingram, 2009; Pagano *et al.*, 2001).

Numerous water management short-term and mid-term decisions are made on the S2S time scale including reservoir operations, water allocation, flood control, hydropower generation, water treatment, and in-stream supported releases (Bolson *et al.*, 2013). Decisions depend largely on streamflow forecasts, many of which are provided by the National Weather Service River Forecasting Centers (RFCs) and National Resource Conservation Service (NRCS) in the United States (US) (T. Pagano et al., 2014). In the Colorado River Basin, a river managed by the

Bureau of Reclamation, streamflow forecasts produced by the Colorado Basin RFC (CBRFC) are used as inputs to operations and planning models which are used for decision making and risk assessment of potential shortage or surplus basin conditions (Bracken, 2011). These streamflow forecasts are not informed by climate forecasts, even though recent work shows benefits (Lehner et al. 2017). Raff et al. (2013) identified enhancements to climate forecasts to meet the needs of water resource managers in the Bureau of Reclamation and US Army Corps of Engineers in a report documenting short-term water management decisions. Water managers interviewed in the report emphasized the need for better understanding of the skill and reliability of climate forecast products, easily accessible products on different time scales, and products presented in a format easily accessible by operators.

Recently, dynamical climate forecasts generated using initialized global climate models (GCMs) have shown skill improvements at the S2S time scale. One of these dynamical models, the fully coupled atmosphere–ocean–land model Climate Forecast System version 2 (CFSv2), which is run at the National Centers for Environmental Prediction (NCEP), demonstrates skill in projecting climate variables at various leads and seasons over the US and improves upon its predecessor CFSv1 (Saha *et al.*, 2014; Tian *et al.*, 2017; Yuan *et al.*, 2011). Multi-model climate forecast ensembles have also demonstrated improved skill over single models (Becker *et al.*, 2014; Doblas-Reyes *et al.*, 2005; Hagedorn *et al.*, 2005). The North American Multi-model Ensemble (NMME) is an operational seasonal climate forecast system that includes ensemble forecasts (for climate and land surface variables) from seven GCMs, leading to more skillful seasonal climate predictions than from any individual GCM (Becker and van den Dool, 2016; Kirtman *et al.*, 2014; Slater *et al.*, 2016).

These climate model forecasts and verifications are normally presented at a system grid resolution or on North American-wide maps, or for all forecast initializations and lead times, rather than particular seasons that are not easily related to local watershed scales. From a water management perspective, however, climate forecast utility is highly specific to location, time of year, and predictand (Wood *et al.*, 2016). There is a gap between the type of verification, data, and product diagnostics provided by forecast production centers and the skill information most readily interpretable and usable by the water community (Wood and Werner, 2011).

A number of the studies referenced above have explored S2S climate forecast skill, but more can be done to support water managers in incorporating climate forecasts into decision making. Some studies have attempted to address these issues by presenting seasonal climate forecasts on a different spatial scale than the typical gridded scale and by displaying skill metrics that are useful to water managers. Hartmann et al. (2002) explored a framework for evaluating seasonal temperature and precipitation projection performance with metrics more easily digestible by users. The metrics were displayed on the 344 Climate Divisions specified by NOAA's Climate Prediction Center (CPC). Although this spatial scale can be useful, the Climate Divisions are not, by design, aligned with hydrologic boundaries that may be relevant for areas of water manager responsibility. More recently, Bolinger et al. (2017) explored the use of a web-based tool to provide monthly updated water-level projections informed by NMME forecasts in the Great Lakes region. The tool allows users to look at individual NMME model results and probabilities of hydrologic variables for specific regions. It represents an example of a regional water group processing climate outlooks onto spatial scales of interest, which underscores the need to develop a centralized, nationwide system to achieve similar goals.

With this motivation in mind, we present work to address some of the hurdles confronting the widespread use of S2S climate predictions in water management applications, and to bridge the gap for potential stakeholders by enhancing the quality, specificity, and accessibility of S2S predictions. To make S2S prediction more usable, this project aligns climate forecasts with users space-time needs, present data in real-time in user friendly formats (such as CSV files by watershed area), remove systematic climatology biases in forecast products, and produce verification information that is relevant to water sector users.

This paper describes a new real-time experimental effort to develop and demonstrate climate forecasts tailored to water managers by presenting real-time forecasts and verification on a watershed scale over the conterminous United States (CONUS) domain. The effort contributes to a sequence of milestones required to transition research toward implementation in an agency operational center such as CPC. For prototyping and demonstration purposes, the efforts adopts the United States Geological Survey (USGS) hydrologic unit code 4 (HUC-4) delineation, which includes of 202 watersheds, which is a suitable spatial scale to show meaningful variability in climate forecasts, given the de-correlation length scales of common climate variables. In Section 2, we describe reforecast and real-time CFSv2 and NMME forecasts, and forcing datasets. Data processing, verification, and basic bias-correction methodologies for precipitation and temperature reforecasts at bi-weekly, monthly, and seasonal time steps are presented in Section

3. Results from watershed-scale verification are evaluated in Section 4, followed by a discussion of water sector responses to the new products and possible improvements to the S2S watershed climate forecasting system in Section 5.

DATA SOURCES AND PROCESSING

CFSv2 Climate Forecasts

The leading operational S2S climate forecast dataset in the US is generated by CFSv2, a fully coupled atmosphere-ocean-land operational model (Saha *et al.*, 2014). CFSv2 forecasts of temperature and precipitation rate are supported by a separate S2S-scale reforecast dataset, which has a 100 km (0.93 degree) grid resolution at a 6-hour time step from 1999 through 2010. The reforecasts were initialized each day at four synoptic times: 0000 UTC, 0006 UTC, 0012 UTC, and 0018 UTC. The 0000 UTC forecast extends to the end of a full season (end of the fourth month), while the 0006, 0012, and 0018 UTC forecasts extend for 45 days. Less frequent CFSv2 reforecasts, not used in this work, extend to 9 months lead time.

For this work, the raw CFSv2 temperature and precipitation reforecasts were re-projected from a native Gaussian grid to a 1/2th-degree grid, temporally averaged to a daily time step, and areally averaged to USGS HUC-4 spatial units through spatially conservative remapping. Figure 1 displays the 202 HUC-4 watersheds in the CONUS domain. Daily ensemble means were calculated for CFSv2 reforecasts and were temporally averaged to bi-weekly time periods (e.g. 1-2 week, 2-3 week, 3-4 week) to support a skill analysis on the sub-seasonal scale. Climatologies for each watershed, lead, and day of year (DOY) are based on a 15-day window (+/- 7 days from forecasted date). CFSv2 data were obtained online from the NOAA National Center for Environmental Information.

Real-time CFSv2 forecast are initialized each day at the four synoptic times, but in contrast to the retrospective runs, each initialization produces four ensemble members for a total of 16 forecasts each day of various lengths: four extend out to 9 months, three to 1 season, and nine to 45 days. The CFSv2 operational 16 member ensemble is downloaded each day and processed similarly to the reforecasts.

NMME Climate Forecasts

The NMME Phase 2 is a combination of seven global climate models which predict precipitation and temperature (among other variables) at a monthly time step for leads up to 7 months (Kirtman *et al.*, 2014). Reforecasts are available for 1982 to 2010 and real-time model forecasts are available for 2011 to present. The models included in NMME are summarized in Table 1. For more information about each model in NMME, see Kirtman et al. (2014) or Slater et al. (2016), but note that the models included in the NMME have changed over time.

[INSERT TABLE 1 HERE]

Raw temperature and precipitation reforecasts are re-projected from a 1-degree grid onto a 1/2th-degree grid and spatially averaged to HUC-4 spatial units using the same method as CFSv2. The NMME forecast ensemble mean, which is used in calculating several of the evaluation metrics, is calculated by equally weighting each model's ensemble average. A seasonal forecast is calculated by temporally averaging the first three months for the forecast for each model. Climatologies are then established for each NMME model, watershed, and forecasted month or season. Real-time NMME forecasts are updated monthly by the 8th day of each month. The ensembles for each of the 7 models are downloaded and processed to watershed scale monthly. Reforecasts were downloaded from the Climate Prediction Center's website and real-time forecasts are downloaded for the IRI Data Library.

NLDAS Climate Observations

The observational data for this study are derived from Phase 2 of the North American Land Data Assimilation System (NLDAS; Xia et al, 2012). NLDAS data are available at 1/8th-degree grid spacing from 1979 to the present at an hourly temporal resolution. Similar to the CFSv2 reforecasts, NLDAS precipitation and temperature data were spatially and temporally aggregated to a daily time step on a 1/2th-degree grid (common to both datasets) before further aggregation to the sub-seasonal HUC-4 space-time resolution to match CFSv2 and NMME time scales. The choice to move to common grid spacing was for ease of analysis and to reduce disk space used during data processing. NLDAS data were obtained from NASA's Earth Science Data Systems Program websites.

METHODS

Post-processing of Climate Forecasts to Reduce Systematic Biases

Raw GCM forecasts require post-processing due to systematic biases, unreliable ensemble spread, and forecasts not being skillful. Post-processing can take the form of statistical or downscaling to improve the raw output of GCMs. In this project, raw CFSv2 forecasts were bias-corrected using the Quantile Mapping (QM) method. QM removes systematic bias between the forecasted and observed climatologies, but does not further calibrate the forecasts to improve their skill. QM is a general method that has long been applied to weather forecasts (Panofsky and Brier, 1968) and later to climate forecasts (Wood et al., 2002). QM is an effective approach to removing bias, but does not address forecast deficiencies in attributes such as reliability and correlation skill (in some cases QM reduces the skill of the forecast). The distinction between bias-correction and probabilistic forecast calibration is further described in Wood and Schaake (2008) and Zhao et al. (2017), and the effectiveness of QM for post-processing climate model outputs for various applications, including extremes projection, is discussed in Ning et al. (2015), Cannon et al. (2015) and Maraun (2013).

When applied to CFSv2, QM replaces the forecast value with a value from the observed climatology (NLDAS) that has the same quantile. This is done by estimating a pair of cumulative distribution functions (CDFs) for the CFSv2 reforecasts and NLDAS data for each variable, lead, watershed, and time scale (climatologies based on 15-day window of +/- 7 days from forecast date). When forecasted values lie outside the quantile range, the two closest quantiles are used to linearly extrapolate the new value. While QM corrects systematic biases in the first and second moments of the climate forecast distribution, concerns have been raised in various studies about its ability to preserve the extreme indices in the observed distribution and to preserve future trends, and it also does not guarantee that biases will be eliminated for durations not explicitly addressed in the CDF mapping. Nonetheless, it serves well as a first step to addressing major bias-related deficiencies in climate model forecast outputs.

Production of Real-time Web-based S2S Climate Outlooks

After spatial remapping and bias correction, prototype S2S climate data products – forecasts and associated skill analyses – are operationally disseminated by the National Center for Atmospheric Research (NCAR) on a public website to facilitate further product development through interactions with water managers. The website (http://hydro.rap.ucar.edu/s2s/) was built in R using the R package Shiny, which supports the staging of websites that link data and geospatial mapping. Climate products on the website include CFSv2-based bi-weekly climate forecasts for HUC-4 watersheds, and NMME-based monthly and seasonal prediction products. The workflow for product generation is summarized in Figure 2 and described in the previous sections. Raw and bias-corrected CFSv2 products are updated daily on the site and NMME products are updated once per month, when NMME forecast outputs are updated.

Forecast Verification

The anomaly correlation coefficient (ACC) metric is widely used in the climate prediction community to measure the degree of association between the forecast mean and the observations. The square of the ACC represents the fraction of climatological variance (uncertainty) explained by the forecast, where a score of 1 indicates that it provides perfect information and a score of zero means the forecast contains no information. For the purposes of prototyping, the ACC was used here to quantify the skill of the forecasts by calculating the correlation between the reforecasts and observations (or forcing data), as follows (Murphy and Epstein, 1989):

$$ACC = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} - \sqrt{n\sum y^2 - (\sum y)^2}} \quad (1)$$

where x is CFSv2 or NMME reforecast anomalies for each watershed and lead of temperature or precipitation and y is NLDAS anomalies for the same variable, watershed and lead, n is the number of forecasts, and ACC is the anomaly correlation coefficient for the reforecasts and forcing data. Anomalies for CFSv2, NMME, and NLDAS for S2S time scales were calculated using the climatologies described in the previous section.

We also calculate other standard deterministic forecast quality metrics that are familiar to water managers, including forecast bias (i.e. the mean error as a percent of observations for precipitation and as a difference from observations for temperature), and mean absolute error, and we plan to assess probabilistic metrics in the future. Forecast ‘skill’ is a multi-faceted concept, generally reflecting the quality of the forecast as described by various dimensions of forecast performance, such as reliability, discrimination, resolution, error, accuracy, correlation and bias. For the demonstration purposes of this paper, we discuss only the ACC and bias.

RESULTS

The raw and bias-corrected real-time climate forecast products being staged on the website are complemented by maps showing skill metrics for different products, seasons and lead times, which we summarize here. The anomaly correlation coefficient for CFSv2 bi-weekly forecasts (Figure 3), shows that temperature has high skill for the first two bi-weekly periods, especially for weeks 1-2. The skill tends to be lower in the western half of the CONUS domain. By weeks 3-4, there are areas with skill exceeding a 'usability' threshold used by the CPC of ACC = 0.3 along the Atlantic and Gulf coasts but the rest of the domain has very low to no skill (O'Lenic *et al.*, 2008). Precipitation forecasts have high skill (reaching values of 0.72) in the first bi-weekly period, especially on the west coast. Skill drops off significantly for weeks 2-3, especially in the central and eastern CONUS domain, and by weeks 3-4, the forecast has negligible skill.

The climate forecast skill varies considerably depending on the season. Figure 4 depicts CFSv2 weeks 2-3 anomaly correlation of precipitation forecast for four seasons. The west coast, especially watersheds in southern Arizona, and the Midwestern US have the highest skill in the December-February (DJF) season. In March-May (MAM) season, the skill is not as high, but the spatial pattern doesn't vary significantly compared to DJF. During the June-August (JJA) period, the pattern shifts and the watersheds in Nevada and Idaho have the highest skill while the remainder of the CONUS domain has lower. In the September-November (SON) period, the region of highest skill shifts to the southeastern US. The maps in Figure 4 display different patterns of forecast skill compared to the corresponding map in Figure 3. This seasonal dependence on skill over the CONUS domain is apparent for all other leads and variables (not shown here).

NMME monthly anomaly correlation for mean temperature and precipitation are shown in Figure 5. There are three leads shown in the figure which are labeled as months. Month 1 refers to the forecast initialization month, or a lead 0, e.g. for a January NMME forecast, Month 1 would refer to January, Month 2 would be February, and Month 3 would be March. As has been found by other authors (Becker and van den Dool, 2016; Slater *et al.*, 2016), temperature forecasts exhibit skill in month 1, especially in the north central US, but this skill drops off significantly in months 2 and 3. Precipitation has some skill in watershed within California and the south east, but other areas of CONUS display low skill. The anomaly correlation of precipitation forecasts in months 2 and 3 are much lower. These trends in skill are highly seasonally dependent; therefore, there may be skill in months 2 and 3 for specific seasons not observed in the annual figures.

A basic skill assessment is presented, but additional analysis into the sources of predictability were not a component of this work. Many other studies have focused of predictability with CFSv2 and NMME. Sources of predictability in the S2S timescale is dependent on the season and lead. Infanti and Kirkman (2016) explored the relationship between ENSO and NMME forecasts of North American precipitation and temperature forecasts. Dirmeyer and Halder (2016) evaluated the sensitivity, variability, and memory of land surface states in CFSv2 and found that soil moisture memory was important in improving forecast skill during spring and summer.

Quantile mapping was used to remove bias from CFSv2 forecasts. The bias prior to quantile mapping is shown in Figure 6. Temperature bias is positive, meaning CFSv2 is over-forecasting temperature compared to NLDAS. The warm bias in temperature appears to grow with lead time. Climate model forecasts are known to drift (i.e. climatologies changing with lead time). To address any drift in bias, the quantile mapping adjustment is performed as a function of lead time. Precipitation exhibits the opposite trend and is mainly under-forecasted, except in a couple watersheds on the west coast and Texas. The spatial patterns in bias do not vary greatly between time periods. Figure 7 illustrates the result of bias-correction and shows that quantile mapping successfully removed bias from the CFSv2 reforecasts.

Bias-correction removes the average bias but does not necessarily improve the forecast skill. While some studies have shown that bias-correction can slightly degrade correlation skill (e.g. Mendoza *et al.*, 2017), here the sample of forecasts used in training the bias-correction does not have this impact (based on cross-validation; results not shown). In Figure 8, the week 2-3 temperature forecast from raw CFSv2 is compared to the QM approach for the Rio Grande-Amistad watershed in southern Texas. The top pair of 1:1 plots show the modeled versus observed forecasts for the raw and QM methods. The raw CFSv2 forecast shows systematic bias as it slightly under-forecasts temperature. The QM approach illustrates the removal of bias as the forecast shift higher and overlaps the 1:1 line. This can also be seen in the time series plot of temperature forecasts and observations for 2000. The QM forecast shifts the forecast up towards the observed temperature throughout the entire year. Other watersheds show similar results of removal of systematic bias where present.

In addition to the issues with capturing extreme events, QM can alter the modeled covariance of temperature and precipitation by QM treating them independently. In downscaling of daily weather data, it is common (and important) to preserve interrelationships between precipitation, temperature, and other fields because there are strong observable relationships linked by synoptic atmospheric dynamics. For instance, wet/precipitating days typically have a compressed temperature range versus clear days. At the sub-seasonal timescale, this covariance is typically weaker. We nonetheless assess the impact of QM on cross-correlations between precipitation and temperature for sub-seasonal bi-weekly CFSv2 predictands for all of the CONUS HUC-4s, in comparison to observations from NLDAS. We find that QM the impact varies by season and lead time. Figure 9 shows these cross-correlations for NLDAS, and CFSv2 forecasts before and after QM, with the HUC-4s in each subplot sorted from low to high values for observed correlation (with samples sizes for each statistic between X and Y of ~360). QM does not significantly affect cross-correlation for January or July forecasts, but has a larger impact, and one that brings cross-correlations of the CFSv2 forecasts into closer agreement with observations for the April and October forecasts. The disagreement between raw CFSv2 and NLDAS grows slightly with lead time. These results suggest that treating temperature and precipitation independently may be acceptable when using QM at the sub-seasonal timescale, and may even improve cross-correlations where the model is biased relative to the observations.

All results shown above are displayed on the S2S Climate Outlooks for Watersheds web-based tool. The results from the verification assessment on an annual and seasonal basis are displayed in tabs for each climate model. Real-time climate forecasts are available as shown in Figure 10. The tool allows the user to choose the lead, variable, and forecast displayed. They can hover over watersheds to view the forecasted anomaly and choose to view the raw or the bias-corrected output. This allows users to view their specific watersheds forecast as well as verification.

DISCUSSION AND CONCLUSION

The new watershed-scale S2S Climate Outlooks for Watersheds web-based tool offers a new medium for water managers to use climate products. Many academic studies and reports from within the industry indicate that S2S forecasts are obtained and assessed qualitatively by water managers, adding to situational awareness, but are less widely used quantitatively, as data streams input to water management decision support tools and models. Water managers cite perceived poor forecast reliability and skill, mismatched temporal or spatial scales, and lack of resources to ingest forecasts (Bolson *et al.*, 2013; Rayner *et al.*, 2005).

Through this new watershed-scale climate product, we aim to overcome several of these hurdles. The skill and accuracy of climate model forecasts for individual watersheds has been explored using an initial, small set of forecast metrics, and these results are presented in an accessible format. Water managers can use the web-based tool to view real-time forecasts of precipitation and temperature at a bi-weekly, monthly, or seasonal outlook, and work is underway to provide data access to the watershed forecasts (and hindcasts) in accessible formats (both text and NetCDF). These products aim to bridge the gap of accessibility, spatial and temporal scale, and perception of unusable skill by allowing water managers to look at the climate forecast for their region as well as the skill of the forecast itself based on an analysis of the model reforecast.

We presented the new S2S Climate Outlooks web-tool to water managers at agencies across the US. The site was presented to Southern Nevada Water Authority (SNWA) in early 2017. SNWA is a wholesale water provider in the Las Vegas region who use reservoir levels at Lake Powell and Lake Mead to plan for their future water management needs. They were interested in how this product could be used to better inform streamflow forecasts in the Colorado River Basin.

Reservoir operators in Reclamation found this tool informative and useful. Operators use forecasts of streamflow quantity and timing to project operations of their reservoirs, and operators in the western US noted stated that this tool could be useful for timing reservoir releases based on projected temperatures during the snowmelt runoff season. Examples of this operation include the Upper Colorado River basin, where releases from Flaming Gorge Reservoir are timed to meet the natural peak in runoff from the Yampa River. Temperature and precipitation S2S forecasts could also be useful for determining reservoir releases when attempting to meet storage targets in early summer when reservoirs are being filled. Reservoir system operators also, however, expressed an interest in a wider variety of forecast products, including time series of past forecasts showing their evolution and agreement with observations, and forecasts of full precipitation and temperature fields rather than forecast anomalies. A California-based watershed manager requested the addition finer scale watershed breakdowns for the climate forecasts.

This work would benefit from a structured analysis of user utility but this focus was not part of this work, which was to demonstrate the concept of watershed-based climate forecast products. Formal surveys could be conducted with a diverse audience of public and private stakeholders to provide feedback and inform future tool development.

The S2S Climate Outlooks web-tool presents a skill assessment of raw CFSv2 forecasts. We show that CFSv2 temperature reforecasts exhibit significant correlation skill in the first two bi-weekly periods, especially in weeks 1-2, while moderate in weeks 2-3, followed by limited skill in weeks 3-4. CFSv2 precipitation forecast show skill over the CONUS domain in weeks 1-2 and regionally in weeks 2-3. In addition to being lead dependent, skill varies seasonally as exemplified in the analysis CFSv2 precipitation reforecast of weeks 2-3. NMME reforecasts displays skill in Month 1, especially when predicting temperature. Months 2 and 3 show lower skill, especially for precipitation.

In general, one expects that as skill in a forecast improves, the forecast has greater utility, but the actual utility for decision making varies greatly among users due to a range of factors. These include the resilience of their system to forecast busts, the characteristics of their penalty functions for forecast errors, their sensitivity to forecast accuracy in different parts of the forecast distribution (e.g. high or low flows), among others. Thus we can only suggest utility based on climate forecast skill. For example, a water manager on the west coast can have reasonable confidence in the spring (MAM) 2-3 week CFSv2 precipitation forecast, but a water manager in Upper Colorado should not have high confidence in the spring 2-3 week precipitation forecast. Similarly, winter (DJF) NMME forecasts of Month 2 temperature for the Pacific Northwest are skillful, but show limited skill during the summer.

The raw model forecasts also contain substantial biases, and we find that the application of quantile mapping to post-process the CFSv2 successfully removed bias from CFSv2 bi-weekly reforecasts for precipitation and temperature. Quantile mapping removes systematic bias between the forecasts and observations but does not improve skill or alter forecast reliability directly as a forecast calibration method might. To improve the skill of the climate forecasts, further work is underway to develop statistical post-processing procedures on a watershed by watershed scale that harness larger scale circulation patterns, variability and potential predictability.

At present, this paper describes the first steps toward addressing hurdles to widespread use of S2S prediction in water management applications. The S2S Climate Outlooks for Watersheds tool presented here enhances the quality, specificity, and accessibility of S2S climate prediction. With wider use of the web-based tool, we intend to improve the product based on user feedback.

ACKNOWLEDGEMENTS

This study was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections program under award #NA16OAR4310138, and the NOAA Climate Testbed, and also benefitted from participation in the NOAA MAPP Subseasonal to Seasonal Climate Testbed. The project is grateful for high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We also appreciate data assistance and insights from Flavio Lehner and Andrew Newman, as well as Andrea Ray, Joe Barsugli, Peitao Peng, and Dave Dewitt for supportive comments.

LITERATURE CITED

Becker, E., and van den Dool, H., 2016. "Probabilistic Seasonal Forecasts in the North American Multimodel Ensemble: A Baseline Skill Assessment." *Journal of Climate* 29 (8): 3015–3026. https://doi.org/10.1175/JCLI-D-14-00862.1

Becker, E., den Dool, H. van, and 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

Bolinger, R. A., Gronewold, A. D., Kompoltowicz, K., and Fry, L. M., 2017. "Application of the NMME in the Development of a New Regional Seasonal Climate Forecast Tool." *Bulletin of the American Meteorological Society* 98 (3): 555–564. https://doi.org/10.1175/BAMS-D-15-00107.1

Bolson, J., Martinez, C., Breuer, N., Srivastava, P., and Knox, P., 2013. "Climate information use among southeast US water managers: beyond barriers and toward opportunities." *Regional Environmental Change* 13 (1): 141–151. https://doi.org/10.1007/s10113-013-0463-1

Bracken, C. W., 2011. "Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management." MS thesis, University of Colorado Boulder.

Callahan, B., Miles, E., and Fluharty, D., 1999. "Policy implications of climate forecasts for water resources management in the Pacific Northwest." *Policy Sciences* 32 (3): 269–293. https://doi.org/10.1023/A:1004604805647

Cannon, A. J., Sobie, S. R., and Murdock, T. Q., 2015. “Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?” *Journal of Climate* 28 (17): 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Dilling, L., and Lemos, M. C., 2011. “Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy.” *Global Environmental Change* 21 (2): 680–689. https://doi.org/10.1016/j.gloenvcha.2010.11.006

Dirmeyer, P.A. and S. Halder, 2016. “Application of the Land–Atmosphere Coupling Paradigm to the Operational Coupled Forecast System, Version 2 (CFSv2).” *Journal of Hydrometeorology* 18 (1): 85–108. https://doi.org/10.1175/JHM-D-16-0064.1

Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N., 2005. “The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination.” *Tellus A* 57 (3): 234–252. https://doi.org/10.1111/j.1600-0870.2005.00104.x

Feldman, D. L., and Ingram, H. M., 2009. “Making Science Useful to Decision Makers: Climate Forecasts, Water Management, and Knowledge Networks.” *Weather, Climate, and Society* 1 (1): 9–21. https://doi.org/10.1175/2009WCAS1007.1

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N., 2005. “The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept.” *Tellus A* 57 (3): 219–233. https://doi.org/10.1111/j.1600-0870.2005.00103.x

Hartmann, H. C., Pagano, T. C., Sorooshian, S., and Bales, R., 2002. “Confidence Builders: Evaluating Seasonal Climate Forecasts from User Perspectives.” *Bulletin of the American Meteorological Society* 83 (5): 683–698. https://doi.org/10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2

Infanti, J.M. and B.P. Kirtman, 2016. “North American Rainfall and Temperature Prediction Response to the Diversity of ENSO.” *Climate Dynamics* 46 (9–10): 3007–3023. https://doi.org/10.1007/s00382-015-2749-0

Kirchhoff, C. J., Lemos, M. C., and Engle, N. L., 2013. “What influences climate information use in water management? The role of boundary organizations and governance regimes in Brazil and the U.S.” *Environmental Science and Policy* 26, 6–18. https://doi.org/10.1016/j.envsci.2012.07.001

Kirtman, B.P., D. Min, J.M. Infanti, J.L. Kinter, D.A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M.P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D.G. DeWitt, M.K. Tippett, A.G. Barnston, S. Li, A. Rosati, S.D. Schubert, M. Rienecker, M. Suarez, Z.E. Li, J. Marshak, Y.-K. Lim, J. Tribbia, K. Pegion, W.J. Merryfield, B. Denis, and E.F. Wood, 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

Lehner, F., Wahl, E. R., Wood, A. W., Blatchford, D. B., and Llewellyn, D., 2017. "Assessing recent declines in Upper Rio Grande runoff efficiency from a paleoclimate perspective." *Geophysical Research Letters* 44 (9): 4124–4133. https://doi.org/10.1002/2017GL073253

Maraun, D., 2013. "Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue." *Journal of Climate* 26 (6): 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

Mendoza, P.A., A.W. Wood, E. Clark, E. Rothwell, M.P. Clark, B. Nijssen, L.D. Brekke, and J.R. Arnold, 2017. "An Intercomparison of Approaches for Improving Operational Seasonal Streamflow Forecasts." *Hydrology and Earth System Science* 21 (7): 3915–3935. https://doi.org/10.5194/hess-21-3915-2017

Merryfield, W.J., W.-S. Lee, G.J. Boer, V.V. Kharin, J.F. Scinocca, G.M. Flato, R.S. Ajayamohan, J.C. Fyfe, Y. Tang, and S. Polavarapu, 2013. "The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization." *Monthly Weather Review* 141 (8): 2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1

Molod, A., 2012. "The GEOS-5 Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna." *Technical report series on global model data assimilation* 28: 175. https://ntrs.nasa.gov/search.jsp?R=20120011790.

Murphy, A. H., and Epstein, E. S., 1989. "Skill Scores and Correlation Coefficients in Model Verification." *Monthly Weather Review* 117 (3): 572–582. https://doi.org/10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2

Ning, L., Riddle, E. E., and Bradley, R. S., 2015. "Projected Changes in Climate Extremes over the Northeastern United States." *Journal of Climate* 28 (8): 3289–3310. https://doi.org/10.1175/JCLI-D-14-00150.1

O'Lenic, E. A., Unger, D. A., Halpert, M. S., & Pelman, K. S., 2008. "Developments in Operational Long-Range Climate Prediction at CPC." *Weather and Forecasting* 23 (3): 496–515. https://doi.org/10.1175/2007WAF2007042.1

Pagano, T. C., Hartmann, H. C., and Sorooshian, S., 2001. "Using Climate Forecasts for Water Management: Arizona and the 1997–1998 El Niño." *Journal of the American Water Resources Association* 37 (5): 1139–1153. https://doi.org/10.1111/j.1752-1688.2001.tb03628.x

Panofsky, H. A., and Brier, G. W., 1968. "Some applications of statistics to meteorology." University Park: Penn. State University, College of Earth and Mineral Sciences.

Raff, D. A., Brekke, L., Werner, K., Wood, A., and White, K., 2013. "Short-Term Water Management Decisions: User Needs for Improved Climate, Weather, and Hydrologic Information." *Bureau of Reclamation, U.S. Army Corps of Engineers and National Oceanic and Atmospheric Administration, Technical Report CWTS-2013-1*. http://www.ccawwg.us/docs/Short-Term_Water_Management_Decisions_Final_3_Jan_2013.pdf

Rayner, S., Lach, D., and Ingram, H., 2005. "Weather Forecasts are for Wimps: Why Water Resource Managers Do Not Use Climate Forecasts." *Climatic Change* 69 (2–3): 197–227. https://doi.org/10.1007/s10584-005-3148-z

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E., 2014. "The NCEP Climate Forecast System Version 2." *Journal of Climat*, 27 (6): 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Slater, L. J., Villarini, G., and 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* 1–16. https://doi.org/10.1007/s00382-016-3286-1

Tian, D., Wood, E. F., and Yuan, X., 2017. "CFSv2-based sub-seasonal precipitation and temperature forecast skill over the contiguous United States." *Hydrology and Earth System Sciences* 21 (3): 1477–1490. http://dx.doi.org/10.5194/hess-21-1477-2017

Vecchi, G.A., T. Delworth, R. Gudgel, S. Kapnick, A. Rosati, A.T. Wittenberg, F. Zeng, W. Anderson, V. Balaji, K. Dixon, L. Jia, H.-S. Kim, L. Krishnamurthy, R. Msadek, W.F. Stern, S.D. Underwood, G. Villarini, X. Yang, and S. Zhang, 2014. "On the Seasonal Forecasting of

Regional Tropical Cyclone Activity." *Journal of Climate* 27 (21): 7994–8016. https://doi.org/10.1175/JCLI-D-14-00158.1

Vernieres, G., M. Rienecker, R. Kovach, and C. Keppenne, 2012. "The GEOS-IODAS: Description and Evaluation." *Technical Report Series on Global Modeling and Data Assimilation* Volume 30. Goddard Space Flight Center. http://gmao.gsfc.nasa.gov/pubs/docs/Vernieres589.pdf.

White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J.T., Lazo, J.K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A.J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A.P., Eggen, B., Graham, R., Kjellström, E., Becker, E., Pegion, K.V., Holbrook, N.J., McEvoy, D., Depledge, M., Perkins-Kirkpatrick, S., Brown, T.J., Street, R., Jones, L., Remenyi, T.A., Hodgson-Johnston, I., Buontempo, C., Lamb, R., Meinke, H., Arheimer, B., Zebiak, S.E., 2017. "Potential applications of subseasonal-to-seasonal (S2S) predictions." *Meteorological Applications* 24 (3): 315–325. https://doi.org/10.1002/met.1654

Wood, A. W., and Werner, K., 2011. "Development of a Seasonal Climate and Streamflow Forecasting Testbed for the Colorado River Basin." Presented at the 36th NOAA Annual Climate Diagnostics and Prediction Workshop, Fort Worth, TX.

Wood, A. W., and Schaake, J. C., 2008. "Correcting Errors in Streamflow Forecast Ensemble Mean and Spread." *Journal of Hydrometeorology* 9 (1): 132–148. https://doi.org/10.1175/2007JHM862.1

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P., 2002. "Long-range experimental hydrologic forecasting for the eastern United States." *Journal of Geophysical Research: Atmospheres* 107 (D20): 4429. https://doi.org/10.1029/2001JD000659

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., and Clark, M., 2016. "Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill." *Journal of Hydrometeorology* 17 (2): 651–668. https://doi.org/10.1175/JHM-D-14-0213.1

Yuan, X., Wood, E. F., Luo, L., and Pan, M., 2011. "A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction." *Geophysical Research Letters* 30 (13): L13402. https://doi.org/10.1029/2011GL047792

Zhang, S., M.J. Harrison, A. Rosati, and A. Wittenberg, 2007. System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Monthly Weather Review* 135 (10): 3541–3564. https://doi.org/10.1175/MWR3466.1

Zhao, T., Bennett, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E., and Ramos, M.-H., 2017. "How Suitable is Quantile Mapping For Postprocessing GCM Precipitation Forecasts?" *Journal of Climate* 30 (9): 3185–3196. https://doi.org/10.1175/JCLI-D-16-0652.1

TABLES

Table 1. North American Multi-model Ensemble (NMME) models

Model Acronym	Model Name	Reference
CFSv2	NOAA NCEP Climate Forecast System version 2	Saha *et al.*, 2014
NASA_GEOS5	Goddard Earth Observing System version 5	Vernieres *et al.*, 2012; Molod *et al.*, 2012
CCSM4	NCAR/University of Miami Community Climate System Model version 4	Lawrence *et al.*, 2012
GFDL-CM2.1	Geophysical Fluid Dynamics Laboratory's (GFDL's) Climate Model version 2.1	Zhang *et al.*, 2007
GFDL_FLOR-CM2.5	GFDL's Climate Model version 2.5 [FLORa06 and FLORb01]	Vecchi *et al.*, 2014
CanCM3	Third Generation Canadian Coupled Global Climate Model	Merryfield *et al.*, 2013
CanCM4	Fourth Generation Canadian Coupled Global Climate Model	Merryfield *et al.*, 2013

LIST OF FIGURES

Figure 2. Summary of methods for processing of Climate Forecast version 2 (CFSv2) and NMME data, and delivering them to an online dissemination platform.

Figure 3. CFSv2 anomaly correlation at bi-weekly time step for temperature and precipitation at a HUC-4 watershed scale.

Figure 4. Anomaly correlation of CFSv2 2-3 week precipitation reforecast for four seasons. Season acronyms contain the first letter of each month included in the season.

Figure 5. NMME anomaly correlation of monthly time periods for temperature and precipitation at a HUC-4 watershed scale.

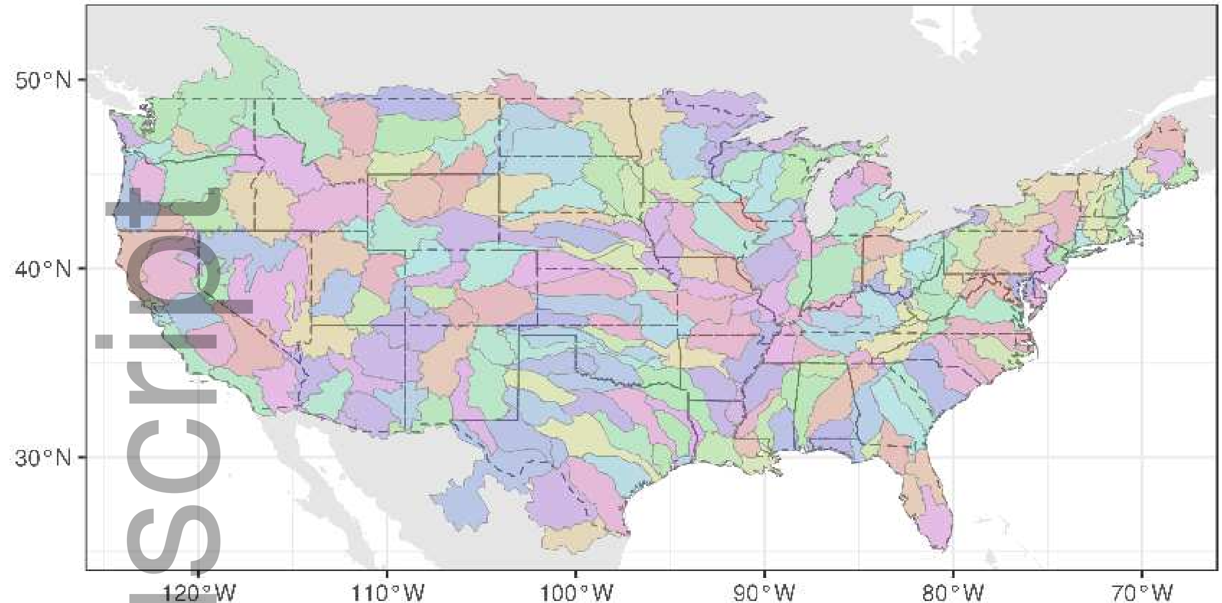
Figure 6. Bias of raw CFSv2 temperature (degrees Centigrade) and precipitation rate (mm/d) for each bi-weekly period.

Figure 7. Bias of quantile mapped (QM) CFSv2 bi-weekly forecasts of temperature (degrees Centigrade) and precipitation rate (mm/d) over the 14-day period.

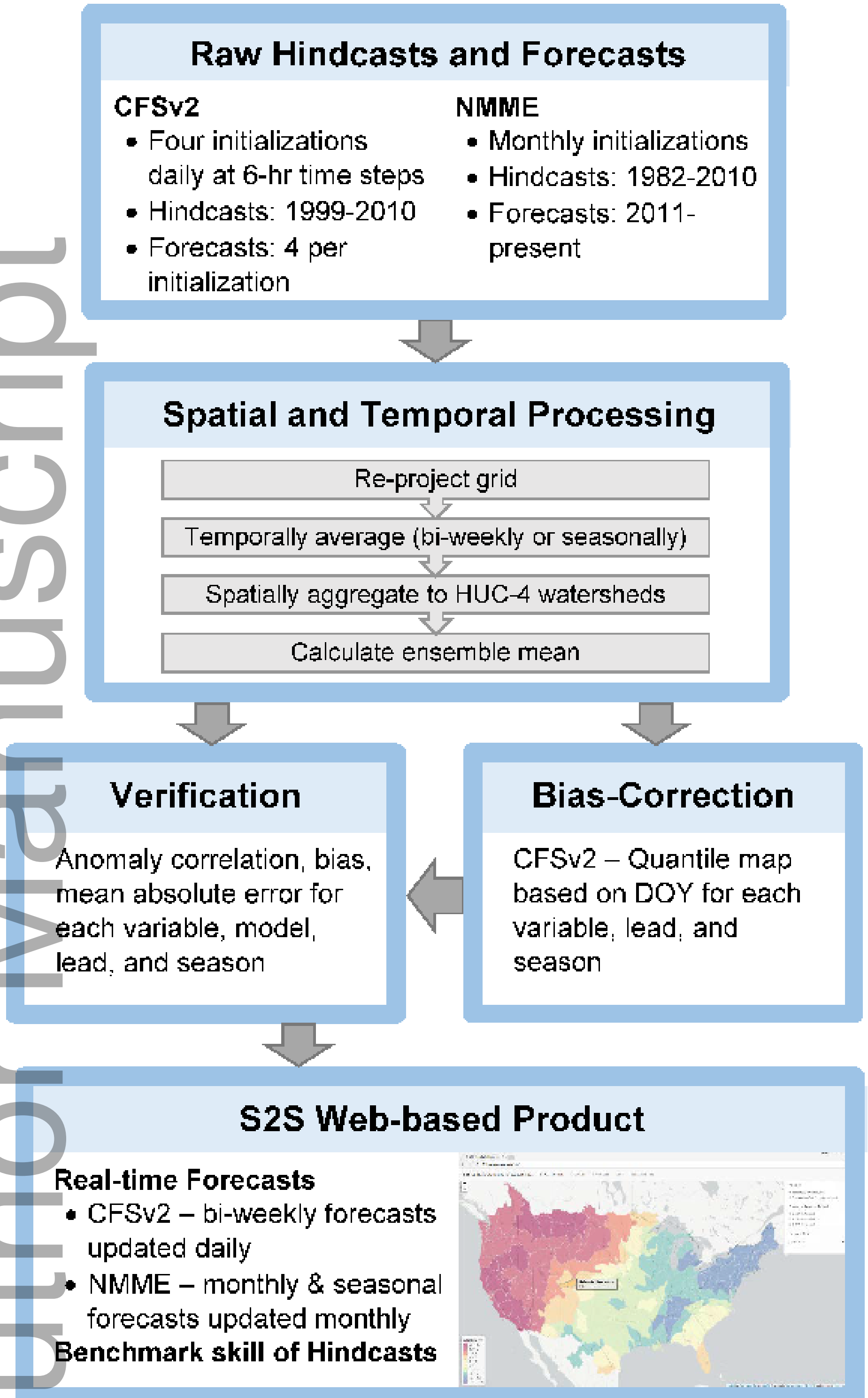
Figure 8. Comparison of Rio Grande-Amistad watershed 2-3 week temperature forecast from CFSv2 and QM. The pair of figures top display the modeled vs. observed forecast for the full time period (1999-2010). The time series plot on the bottom displays the forecast for CFSv2 and QM in comparison to North American Land Data Assimilation System (NLDAS) for 2000.

Figure 9. The bi-weekly cross-correlation of temperature (T) and precipitation (P) for NLDAS, raw CFSv2, and QM CFSv2 for the forecast months January, April, July, and October. The x-axis is the index of the HUC-4s in ascending order of NLDAS cross-correlation.

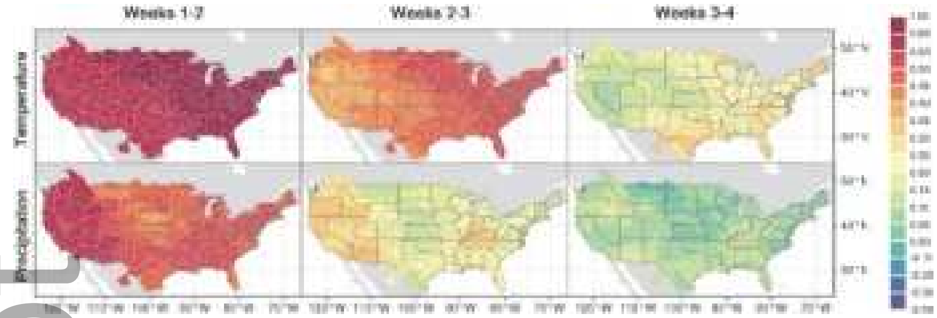
Figure 10. Sub-seasonal to seasonal (S2S) Climate Outlooks for Watersheds web-based tool allows users to look at specific watersheds forecasts and verification metrics.



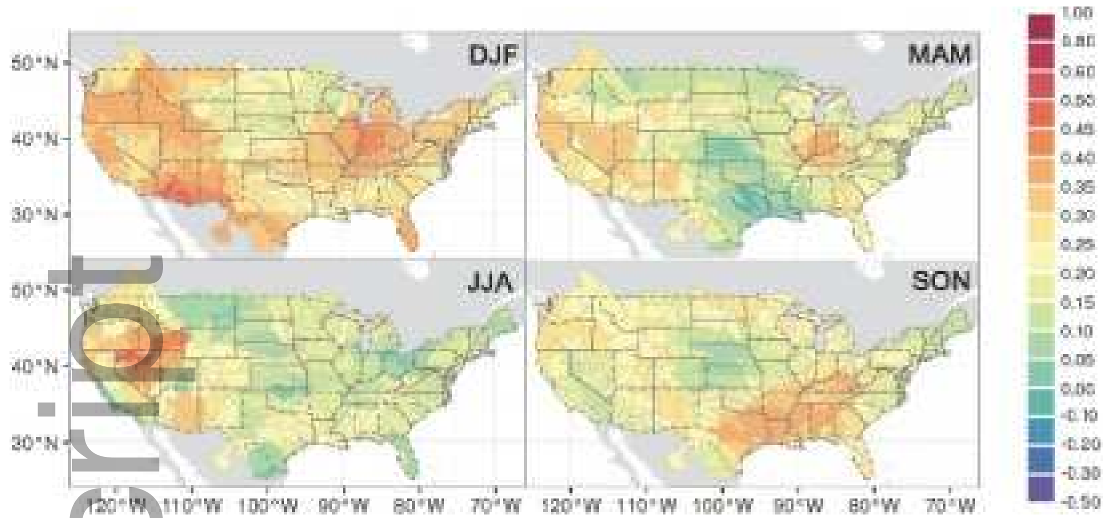
jawra_12746-18-0101_f1.tif



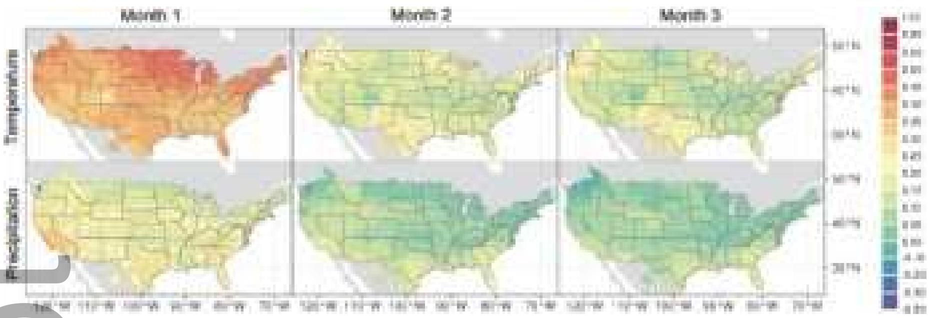
jawra_12746-18-0101_f2.tiff



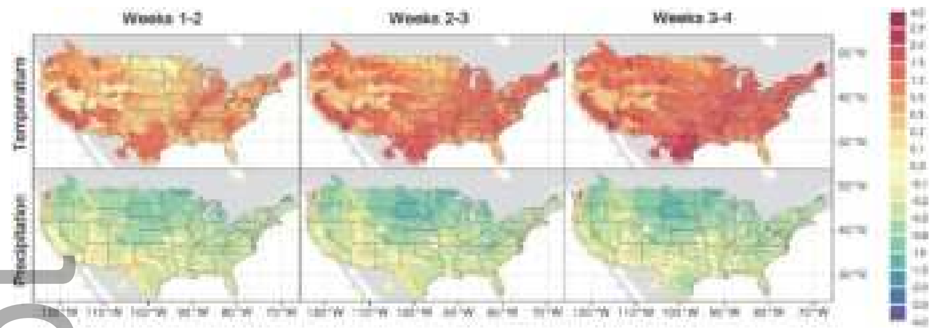
jawra_12746-18-0101_f3.tif



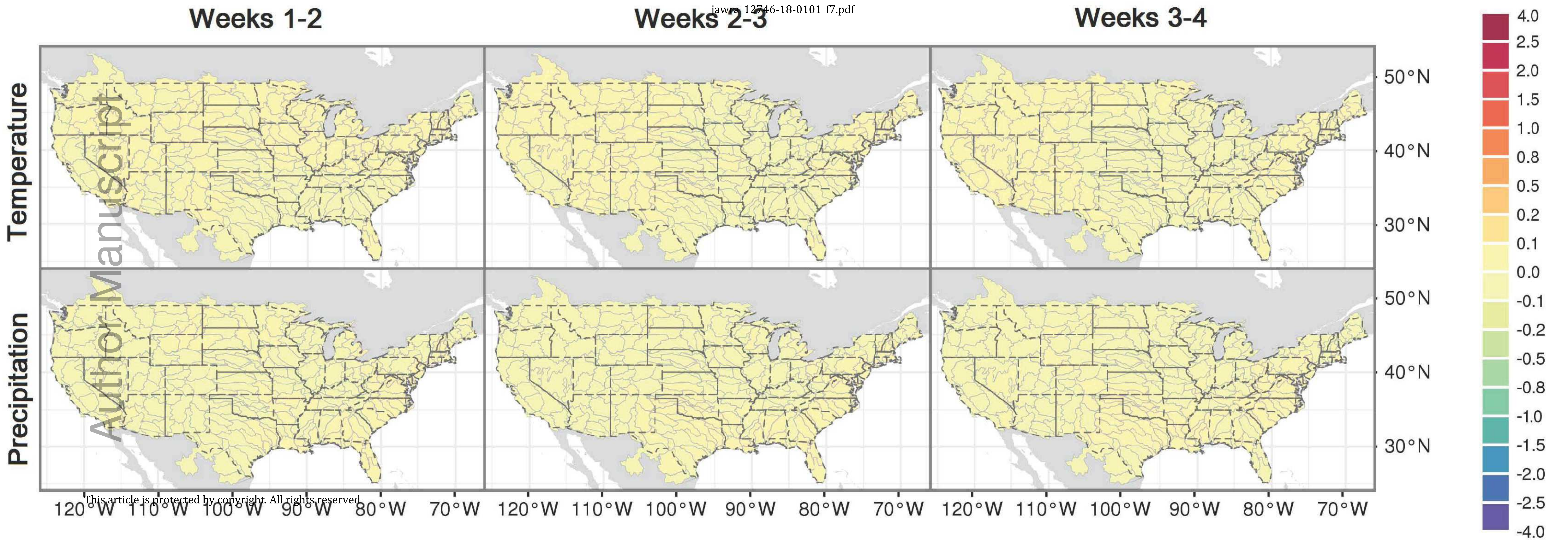
jawra_12746-18-0101_f4.tif

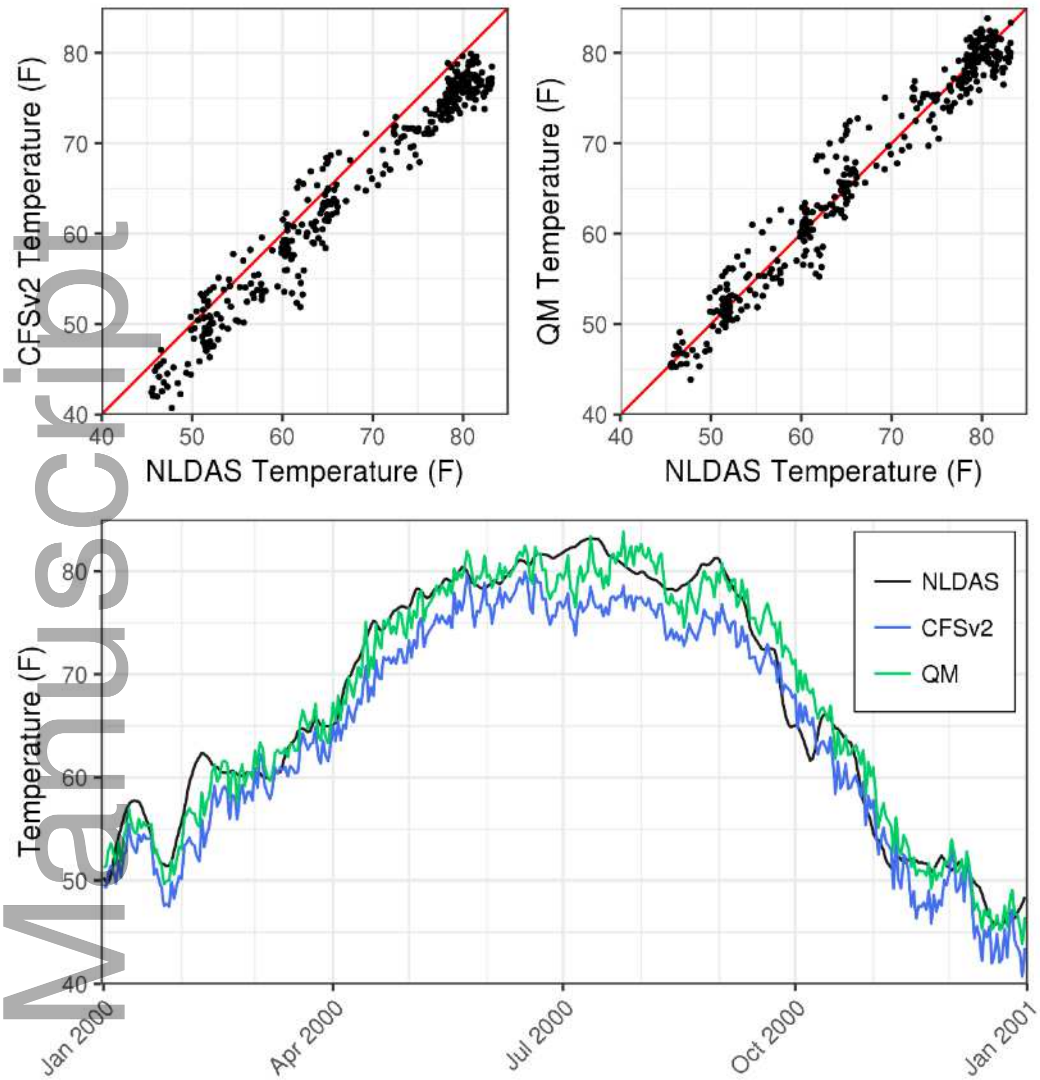


jawra_12746-18-0101_f5.tif

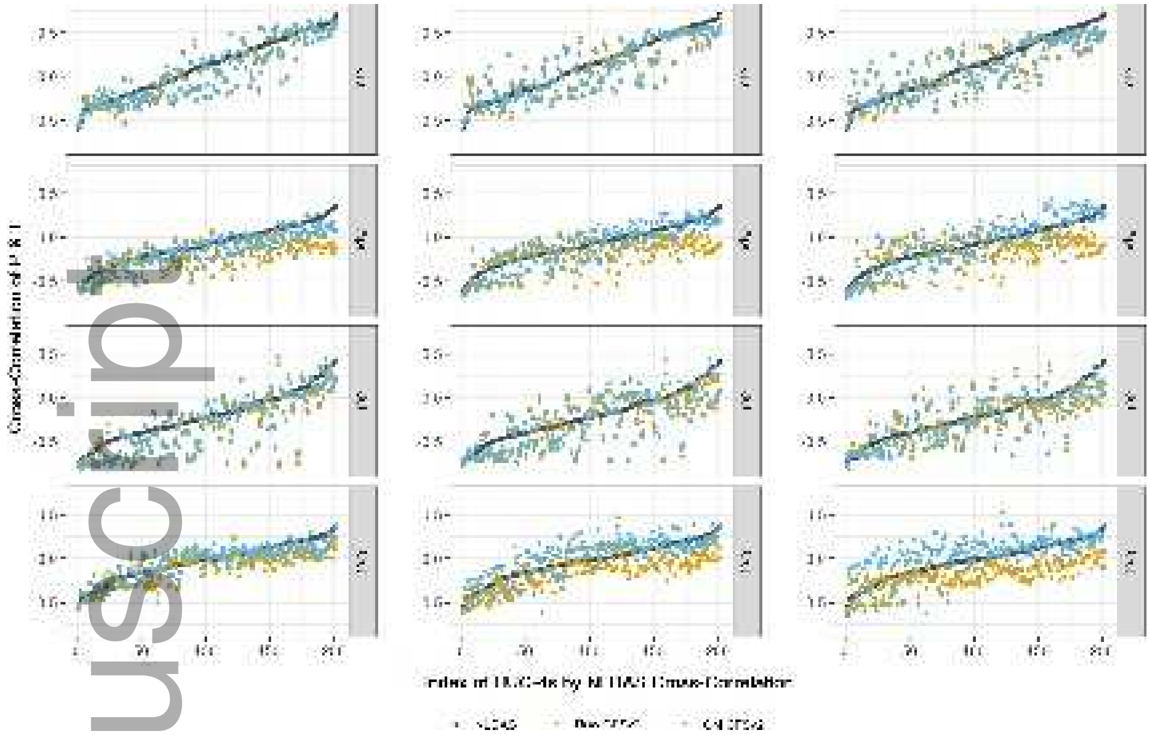


jawra_12746-18-0101_f6.tif

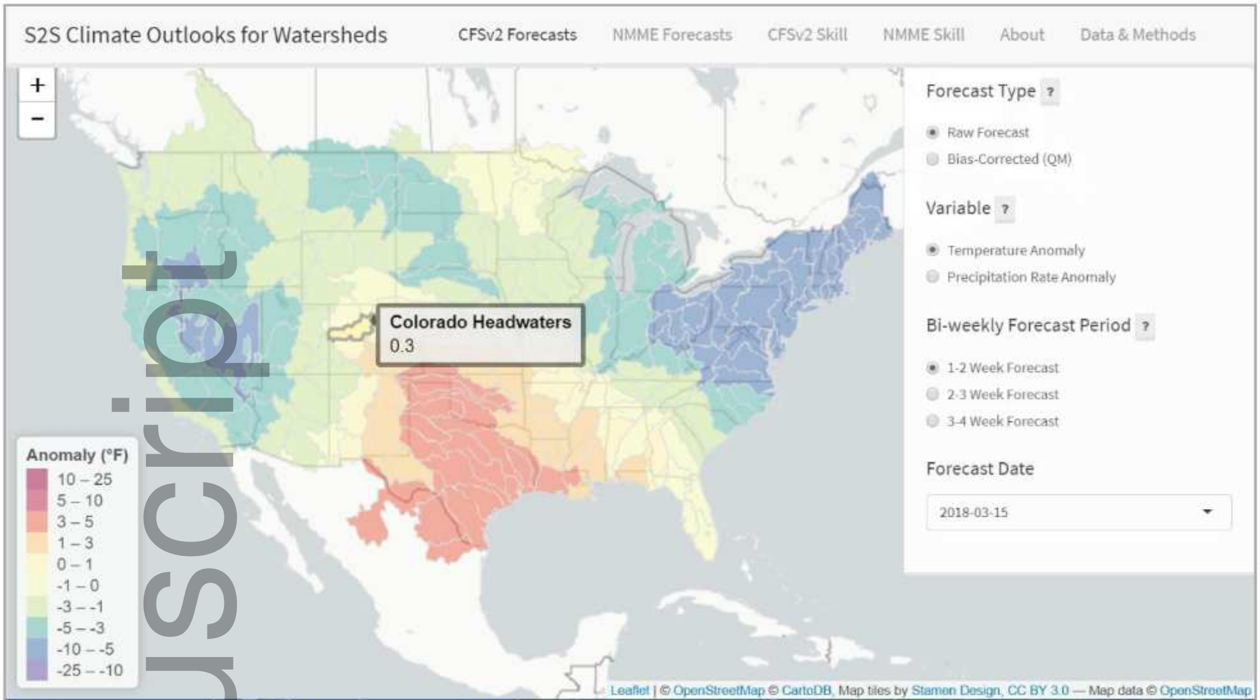

Weeks 1-2
Weeks 2-3
Weeks 3-4
Temperature
Precipitation
50°N
40°N
30°N
50°N
40°N
30°N
120°W
110°W
100°W
90°W
80°W
70°W
4.0
2.5
2.0
1.5
1.0
0.8
0.5
0.2
0.1
0.0
-0.1
-0.2
-0.5
-0.8
-1.0
-1.5
-2.0
-2.5
-4.0



jawra_12746-18-0101_f8.tif



jawra_12746-18-0101_f9.tif



jawra_12746-18-0101_f10.tif