Colorado Airborne Snow Measurement Program:

Water Year 2022 Streamflow Forecast Review

Northern Colorado Water Conservancy District

Colorado Water Plan Grant

June 30th, 2023

Submitted To:

Northern Colorado Water Conservancy District
Attention: Emily Carbone
220 Water Avenue
Berthoud, Colorado 80513

Submitted By:

Lynker Corporation
Graeme Aggett, Principal and
Chief Scientist
5445 Conestoga Court, Suite 100
Boulder, Colorado 80301

Table of Contents

1. **Executive Summary** 7
2. **Introduction** 9
3. **Methods** 11
   - **3.1. Study Sites** 11
   - **3.2. Snow Data Assimilation (DA)** 12
   - **3.3. Historical Water Supply Forecast Evaluation** 13
   - **3.4. SUMMA Model and Hindcasting Approach** 13
4. **Results** 15
   - **4.1. Hydroclimate Data** 15
   - **4.2. Historical Water Supply Forecast Skill** 16
   - **4.3. Experimental Forecast Documentation** 18
   - **4.4. Forecast Error Attribution: Spring Precipitation Anomalies** 21
   - **4.5. SUMMA Hindcast Experiment** 23
5. **Discussion** 25
6. **Conclusions and Recommendations** 27
7. **References** 30
8. **Appendices** 32
   - **8.1. Hydroclimate Data** 32
     - *8.1.1. Taylor River at Taylor Park* 32
     - *8.1.2. Willow Creek at Willow Creek Reservoir* 33
     - *8.1.3. Lake Granby* 34
     - *8.1.4. Dillon Reservoir* 35
     - *8.1.5. Dolores River at Dolores* 36
     - *8.1.6. Fraser River at Winter Park* 37
     - *8.1.7. Conejos River near Mogote* 38
   - **8.2. Historical Water Supply Forecasts** 39
     - *8.2.1. Taylor River at Taylor Park* 39
     - *8.2.2. Willow Creek at Willow Creek Reservoir* 39
     - *8.2.3. Lake Granby* 40
     - *8.2.4. Dillon Reservoir* 40
     - *8.2.5. Dolores River at Dolores* 41
     - *8.2.6. Fraser River at Winter Park* 41
     - *8.2.7. Conejos River near Mogote* 42
   - **8.3. Experimental Forecast Documentation** 43
     - *8.3.1. Water Supply Forecast Figures* 43
     - *8.3.2. WRF-Hydro Experimental Forecasts- Tabulated* 47
     - *8.3.3. CBRFC Experimental Forecasts- Tabulated* 49
   - **8.4. Forecast Error Attribution: Spring Precipitation** 51

8.4.1. Taylor River at Taylor Park *51*
8.4.2. Willow Creek at Willow Creek Reservoir *51*
8.4.3. Lake Granby *52*
8.4.4. Dillon Reservoir *52*
8.4.5. Dolores River at Dolores *53*
8.4.6. Fraser at Winter Park *53*
8.4.7. Conejos River near Mogote *54*
8.5. SUMMA Hindcasts 55
8.5.1. East River at Almont *55*
8.5.2. Taylor River at Taylor Park *56*
8.6. CASM Streamflow Forecast Roundtable Agenda 59
8.7. CASM Streamflow Forecast Roundtable Minutes 60

Figures

Figure 1: CASM pilot basins (blue) from the 2022 program year, as funded by the Colorado Water Conservation Board's Water Plan Grant and matching funds from stakeholders. Each pilot basin was flown by ASO two times between April-May of 2022, with experimental, ASO-informed streamflow forecasts issued by CBRFC and NCAR-RAL for the 2022 snowmelt runoff season. ... 11
Figure 2: SUMMA model of the Taylor River at Taylor Park, Colorado from Dr. Andrew Wood of NCAR-CGD. The SUMMA model shown here is discretized by elevation bands ('Elev'; high and low) and radiation inputs ('Rad'; high and low) into four hydrologic response units (HRUs; colored) for each geographical response units (GRUs; black outline) for a total of 12 HRUs. This model discretization aims to improve snow process simulations using a computationally efficient strategy; it was not implemented here due to the limited scope of this pilot study. ... 14
Figure 3: East River at Almont forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL stations. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL 380 and 737; bottom panel). Temperature and precipitation are measured on a sub-basin level (HUC12 basins, labeled as GRUs, or Geographical Regional Units), as shown by the colored dots, with basin mean conditions denoted by the black line. The East River at Almont consists of ten GRUs, or HUC12 basins. Grey dashed lines show air temperature and precipitation trendlines; black solid lines show a five-year rolling average. ... 16
Figure 4: East River at Almont April 1st forecasts of April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right) in units of thousand-acre-feet (KAF). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, orange dot notes the 50% level (median). Dashed lines represent the 20% bounds around the 1:1 line. ... 17
Figure 5: Official and experimental water supply forecasts for the East River at Almont, showing the median April-July forecasted volumes in 2022 (in thousand acre-feet, or KAF) bounded with error bars by the 10-90% forecast exceedance volumes. 'Forecast 1' (grey) and 'Forecast 2' (purple) issue dates all vary slightly across models. For example, depending on the model, 'Forecast 1' corresponds to either a) the first ASO flight date (here, April 21st, 2022), b) the nearest official forecast date relative to the ASO flight time (e.g., May 1st for the NRCS and CBRFC forecasts in grey). The black horizontal line is the NRCS-reported adjusted April-July volume; the red horizontal dashed lines is the NRCS-reported 1991-2020 median volume. Late-April WRF-Hydro forecasts were post-processed from the provided April-September volumes to April-July volumes by subtracting out NRCS-reported average flows for August and September; see Table 6 and Table 7 in the Appendix for additional information. ... 19
Figure 6: April 1st forecast percent error (for the 50% exceedance forecast of April-July adjusted volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right) for the East River at Almont (2000-2022), as measured by a single-ensemble member of the GMET retrospective forcing dataset. Results show that about 60% of forecast error in the East River is explained by spring precipitation anomalies. ... 22
Figure 7: A figure from CBRFC's Water Year in Review report showing CBRFC-estimated spring precipitation errors (orange) for April 1st forecasts of April-July volumes across four basins in Colorado (NOAA, 2022). TRAC2: Taylor Park, UCRC2: Uncompaghre near Ridgeway, BMDC2: Blue Mesa Reservoir, ALEC2: East River at Almont. Results show that approximately 50% of forecast error is from unknown spring precipitation (orange) at the time of the forecast. Blue bars show error from other sources, such as simulated snowpack and/or model structural errors. ... 22
Figure 8: SUMMA bias-corrected ESP hindcast of the East River at Almont for April 1st, 2022, showing 53 ensemble members from meteorological years 1970-2022 (grey traces) with respect to observed flows from the USGS gage (black). Probability distributions for meteorological years 2000-2021 are highlighted in red (10-90th percentile) where the median (50th percentile) is denoted by the solid red line. The 2022 retrospective model run (which uses actual 2022 forcings) is in blue. ... 23
Figure 9: Multi-decadal SUMMA hindcasts (leave-one out bias corrected-ESP for all April 1st dates) and retrospective simulations for the East River at Almont. Here, the top panels show the 2000-2022 time series of the one-year April 1st hindcasts; the bottom panel shows the volumetric April-July flows for the probabilistic hindcasts (boxplots) and deterministic retrospective simulation (teal square) against USGS streamflow observations. Statistics are of the median BC-ESP hindcast. ... 24
Figure 10: Modified figure from Mendoza et al. (2017) comparing water supply forecast skill (as measured by the Continuous Ranked Probability Skill Score, CRPSS, where higher scores are better) for hindcasts of April-July runoff into Dworshak Reservoir, Idaho. Results show that early in the water year, we can improve forecasts using

climate information, while closer to April 1st, we can develop the most robust forecasts by using both watershed (e.g., snowpack) and climate (e.g., El Nino Southern Oscillation indices) information as shown by the ensemble streamflow prediction trace weighting scheme (TWS) in blue. 26
Figure 11: Taylor River at Taylor Park forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 32
Figure 12: Willow Creek forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 33
Figure 13: Lake Granby forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 34
Figure 14: Dillon Reservoir forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 35
Figure 15: Dolores River forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). The dip in mid-1980s air temperature is likely an artifact of a new station coming online, which affects the underlying dataset. 36
Figure 16: Fraser River at Winter Park forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 37
Figure 17: Conejos River near Mogote forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). 38
Figure 18: Taylor River at Taylor Park April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. 39
Figure 19: Willow Creek Reservoir April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. 39
Figure 20: Lake Granby April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. 40
Figure 21: Dillon Reservoir April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. 40
Figure 22: Dolores River at Dolores April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. 41
Figure 23: Fraser River at Winter Park April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from CBRFC. Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. The period of record is limited by the lack of observed data for the Jim Creek Diversion from CO DWR, which is used to calculate the adjusted April-July volume of the Fraser River at Winter Park. 41
Figure 24: Conejos River near Mogote April 1st forecasts for April-September (AMJJAS) volumes vs. observed AMJJAS streamflow from NRCS. Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the

observed volume. WGRFC does forecast for this point, however the period of record is limited and thus is excluded from the study. ... 42
Figure 25: Taylor Reservoir official and experimental water supply forecasts. SUMMA hindcasts are not available for Taylor Reservoir, since the model was run for the Taylor River at Taylor Park forecast point just upstream of the reservoir. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing... 43
Figure 26: Willow Creek at Willow Creek Reservoir official and experimental water supply forecasts. Forecast 1 from WRF-Hydro was not issued due to technical issues of the forecasting system. ... 43
Figure 27: Lake Granby official and experimental water supply forecasts. WRF-Hydro forecasts were issued for naturalized flows, which poorly represented the NRCS-reported 2022 AMJJ adjusted volumes. Additionally, Forecast 1 from WRF-Hydro was not issued due to technical issues of the forecasting system. ... 44
Figure 28: Dillon Reservoir official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing. ... 44
Figure 29: Dolores River at Dolores official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing. ... 45
Figure 30: Fraser River at Winter Park official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing. ... 45
Figure 31: Conejos River at Mogote official and experimental water supply forecasts; note that the Conejos River forecast target period is April-September, not April-July. ... 46
Figure 32: An example of CBRFC water supply forecasts for 2022, where official forecast guidance is in purple compared with ASO-informed forecasts in light blue (April 21st, 2022, and May 18th, 2022)... 49
Figure 33: Taylor River at Taylor Park April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset. ... 51
Figure 34: Willow Creek Reservoir April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset. ... 51
Figure 35: Lake Granby April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset. 52
Figure 36: Dillon Reservoir April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset... 52
Figure 37: Dolores River at Dolores April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset. ... 53
Figure 38: Fraser River at Winter Park April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for CBRFC (right), as measured by the GMET forcing dataset. ... 53
Figure 39: Conejos River near Mogote April 1st forecast percent error (for the 50% exceedance forecast of April-September volumes) vs. April-September precipitation anomalies for NRCS, as measured by the GMET forcing dataset... 54
Figure 40: SUMMA raw (top) and bias-corrected (bottom) ESP hindcast of the Taylor River at Taylor Park for April 1st, 2022, showing 53 ensemble members from meteorological years 1970-2022 (grey traces) with respect to observed flows from the USGS gage (black). Probability distributions for meteorological years 2000-2021 are highlighted in red (10-90th percentile) where the median (50th percentile) is denoted by the solid red line. The 2022 retrospective model run (i.e., using actual 2022 forcings) is in blue. All simulations are bias-corrected using a Leave One-Out (LOO) approach. ... 56
Figure 41: Multi-decadal SUMMA hindcasts (bias corrected-ESP for all April 1st dates) and retrospective simulations for the Taylor River at Taylor Park. Here, the top panels show the 2000-2022 time series; the bottom panel shows the volumetric April-July flows for the probabilistic hindcasts and deterministic retrospective simulation against USGS observations. Statistics are of the median BC-ESP hindcast. ... 58

Tables

Table 1: Study sites and available water supply forecasts from official (CBRFC, NRCS) and experimental (CBRFC, NCAR-RAL's WRF-Hydro, and NCAR-CGD'S SUMMA) modeling systems. Adjusted streamflow volumes are pulled from NRCS historical data, with the exception of the Fraser River at Winter Park, which was calculated from gaged streamflow and diversions. 12
Table 2: Basin mean hydroclimate statistics from the GMET retrospective forcing dataset (air temperature and precipitation) and the nearest NRCS SNOTEL station with the longest record (peak SWE) for the period 1970-2022 (where data are available). 15
Table 3: April 1st water supply forecast skill from NOAA's RFCs and USDA's NRCS for the period 2000-2022, as measured by the correlation coefficient (r^2), mean absolute error (MAE; in thousand-acre feet; KAF), and percent bias (PBIAS), relative to the 50% exceedance value of the forecasts. Basin average statistics are presented for the normalized error metrics (i.e., r^2 and PBIAS). 17
Table 4: Water supply forecast error (right of the black bar, as a percent of NRCS-reported April-July adjusted volumes) for the 50% exceedance value from the April 1st forecast/hindcast (NRCS, CBRFC) or the nearest first ASO flight (CBRFC, WRF-Hydro; generally in late April). Note that the ASO-informed forecasts from CBRFC and NCAR-RAL were issued towards late April (when ASO was flown), which is nearly a month after the NRCS and CBRFC April 1st forecasts. 20
Table 5: Spring (April-July) precipitation anomalies as measured by the retrospective GMET forcing dataset and correlation coefficients with forecast percent errors for both NRCS and CBRFC April 1st forecasts for the period 2000-2022. 21
Table 6: WRF-Hydro experimental water supply forecasts, flight one. All volumes are in KAF. 47
Table 7: WRF-Hydro experimental water supply forecasts, flight two. All volumes are in KAF. 48
Table 8: CBRFC experimental (ASO SWE-data assimilation) water supply forecasts, flight one. All volumes are in KAF. 49
Table 9: CBRFC experimental (ASO SWE-data assimilation) water supply forecasts, flight two. All volumes are in KAF. 50
Table 10: SUMMA simulation statistics for April 1st to September 30th, 2022, East River at Almont. 55
Table 11: SUMMA simulation statistics for April 1st to September 30th, 2022, Taylor River at Taylor Park. 57

1. Executive Summary

In 2021, the Colorado Water Conservation Board (CWCB), together with key stakeholders, funded a pilot study to evaluate the benefits of airborne lidar for improved snow monitoring and streamflow forecasting across important headwater basins in Colorado. Under this CWCB Water Plan Grant (WPG), the Colorado Airborne Snow Measurement (CASM) program planning team, which includes the WPG awardee, Northern Colorado Water Conservation District (Northern Water) along with Denver Water, Dolores Conservation District, Saint Vrain & Left Hand Water Conservancy District, Colorado River District, and the Colorado Water Conservation Board, worked closely with the Airborne Snow Observatory (ASO) Inc. to coordinate and fly a total of 16 spring snow lidar surveys across eight pilot basins in Colorado between April and May 2022. These ASO snow surveys were assimilated into two real-time experimental streamflow forecasting systems with the goal of evaluating the utility of these data in a seasonal water supply forecasting framework. In support of this effort, Lynker was tasked with reviewing these 2022 streamflow forecasts, facilitating a streamflow forecasting roundtable among modelers and practitioners, and assessing the CASM streamflow forecasting program more broadly, including developing a set of recommendations to support the improved use of ASO snow data in hydrologic models across Colorado. This report summarizes the findings of our study.

With limited high-resolution ASO snow data in Colorado to date, there is at this point no evidence to show that the inclusion of spatially distributed snow observations markedly and consistently improves simulations of streamflow in the Rocky Mountains. In the Colorado Basin River Forecast Center (CBRFC)'s "Water Year in Review" (NOAA, 2022), CBRFC noted that while the assimilation of high-resolution snowpack data into hydrologic models provides a unique research opportunity, CBRFC experimental forecasts that use the direct insertion of ASO's snowpack estimates with CBRFC's existing operational hydrologic models have demonstrated inconsistent results in Colorado. While this analysis was only from a single year (2022), these findings highlight the challenge for modelers in translating improved snow data products into improved streamflow forecasts. Furthermore, it highlights the importance of supplemental CASM tasks with the express objective of evaluating multiple modeling approaches and their ability to better integrate ASO snow data and better simulate the snow processes so important to Colorado's hydrology. To this end, Lynker led a retrospective streamflow forecast evaluation effort of official and experimental water supply forecasts from CBRFC and the National Center for Atmospheric Research's Research Applications Lab (NCAR-RAL) with a limited retrospective implementation of an additional hydrologic model from NCAR's Climate & Global Dynamics Lab (NCAR-CGD).

One critical objective of this work was to establish a quantitative benchmark against which to measure future CASM-supported experimental streamflow forecasts. In addition to establishing these reference benchmarks, this study also aimed to better characterize historical sources of water supply forecast error to estimate the constraints of possible improvements from snow data assimilation (and other demonstrated methods). Results showed that across the six CASM pilot basins for which the Natural Resource Conservation Service (NRCS) and CBRFC both forecast, average error of April 1st 50% exceedance forecasts of April-July volumes between the years of 2000-2022 was approximately 16.1% for NRCS forecasts and 16.8% for CBRFC forecasts; our empirical analyses estimate that an average of approximately half (but from 30-70%) of this error was related to spring precipitation anomalies, which is consistent with findings from CBRFC. These results suggest that, lacking improved seasonal precipitation forecasts, the average ceiling for water supply forecast improvement in Colorado is on the order of 5-10% of total April-July inflows. While this does not preclude greater improvements in individual years, it does underscore the importance of systematic model hindcast studies (ideally in peer-reviewed publications) that aim to understand historical sources of forecast error as well as how (and to what extent) ASO data assimilation might resolve these errors within our models.

This study also demonstrated the suitability of intermediate-complexity hydrologic models for streamflow forecasting in Colorado through a pilot study of the SUMMA model. In addition to a targeted 2022 hindcast experiment in two CASM basins, we also demonstrated the ability of intermediate-complexity and scale models to be used in multi-decadal hindcasting, or re-forecasting, studies. These hindcasting studies (the first by a CASM-supported project) are critical for demonstrating long-term model performance, but also understanding any limitations of the modeling framework. The SUMMA re-forecasts for both the East and Taylor Rivers were, by all measures, quite skillful: for the 22 hindcast years of this study (2000-2022), correlation coefficient values (r^2) of the 50% exceedance for April-July volumes were ~0.8, with an absolute bias of <3% and a mean absolute error less than or equal to the real-time operational forecasts from CBRFC and NRCS. Increased model spatial resolution (for optimized snow representation) run with high-resolution model forcings along with the assimilation of ASO snowpack data and/or seasonal climate information (all of which are existing capabilities of SUMMA and other hydrologic models) would likely improve upon these performance measures.

The results of Lynker's study highlight how future CASM streamflow forecast improvement efforts would be well-served to focus on evidence-based approaches that allow for direct comparison (benchmarking) with official water supply forecasts, and other experimental approaches, including the use of various data assimilation strategies. This is particularly important given that to date, virtually all ASO snow data assimilation studies have focused on California's Sierra Nevada (primarily in a single basin, the Tuolumne River, though increasingly across the California), which is both climatically and hydrologically unique from the Rocky Mountains of Colorado. Evaluating progress in a scientifically robust and defensible way should be a priority for the program if additional years of ASO data become available in Colorado; publishing these results in refereed journals would further these objectives as well as programmatic credibility. In route to an improved streamflow forecast, Lynker recommends that CASM set the following goals for 2024:

1. **Improved and more balanced engagement with stakeholders, including:**
    - Communicating both the opportunities and challenges of snow data assimilation (DA) in hydrologic models, including discussion of additional forecast improvement techniques
    - Addressing the challenges of translating improved snow data products into improved streamflow forecasts, particularly in an operational capacity

2. **Continuation of the annual "CASM Streamflow Forecasting Roundtable" each fall**
    - Comprised of forecasters, modelers, and key CASM stakeholders
    - With an annual forecast verification of CASM-supported experimental forecasting systems

3. **Funding of additional model hindcasting experiments targeting peer-reviewed publication**
    - Designed to objectively measure streamflow forecast improvements over benchmarks
    - But also the evaluation of other modeling approaches (see recommendation #4)

4. **Exploration of additional modeling approaches and techniques, including:**
    - Models:
        - Other intermediate-complexity hydrologic models with snow-DA (beyond direct insertion), hindcasting, and real-time forecasting capabilities (e.g., NCAR-CGD's SUMMA)
        - Machine-learning models (e.g., LSTM from the NextGen NWM, Upstream Tech, etc.)
    - Integration/assimilation of additional information using improved DA techniques, such as:
        - Other datasets: snow albedo, snow covered area, streamflow
        - Seasonal precipitation and temperature forecasts
    - Next-generation modeling capabilities from federal forecasting agencies:
        - USDA NRCS's new AI-based water supply forecasting system, M4
        - NOAA's NextGen National Water Model (NWM), a model-agnostic framework

5. **Standardization of forecasts with existing federal forecast agency's practices, including:**
    - All forecast points, target periods/datasets, and delivery mechanisms for all future real-time forecasts and hindcast experiments

6. **Collaboration with NOAA's new Cooperative Institute for Research to Operations in Hydrology (CIROH)**
    - CIROH is a new $360M cooperative institute focused on improving hydrologic prediction

Together, these efforts will work towards ensuring that discussions with stakeholders around if and how to continue the CASM program are transparent and evidence-based, and that under the premise of a continued CASM program, 1) ASO-informed seasonal streamflow forecasts provide quantifiable performance benefits over currently available forecasts relative to observations; and 2) models, methods, and techniques identified for CASM experimental streamflow forecasts during 2024 and beyond are well-suited for the needs of Colorado stakeholders going forward.

2. Introduction

The Colorado Airborne Snow Measurement Program (CASM) program together with Airborne Snow Observatories (ASO) Inc. is evaluating the assimilation of high-resolution snowpack data into hydrologic models for the improvement of streamflow forecasts in Colorado. However, given the limited availability of ASO data in Colorado, there is not yet evidence to show that the inclusion of spatially distributed snow observations markedly and consistently improves simulated streamflow in Colorado, particularly on the seasonal time scales relevant to water managers and utilities. In fact, preliminary results from NOAA's Colorado Basin River Forecast Center (CBRFC) spring 2022 forecasts have demonstrated both forecast improvement and degradation when directly assimilating ASO snowpack data into their existing operational snow and hydrologic models in Colorado (NOAA, 2022), underscoring the need for additional data collection and future study.

ASO's airborne lidar technology was originally developed at the National Aeronautics and Space Administration's (NASA) Jet Propulsion Lab (JPL) in Pasadena, California prior to the private technology transfer to ASO Inc. Because of this legacy, to date, virtually all research on the use of ASO data in hydrologic models has been focused in California's Sierra Nevada; only recently have ASO snow data been collected on a limited basis in Colorado and Wyoming. While operationalization of ASO data collection in partnership with California's Department of Water Resources has been a role model for CASM, there are important differences between implementation of ASO snow data assimilation in hydrologic models used in California versus Colorado that must be considered. One key difference is California's Mediterranean climate, which has a much drier spring than Colorado, with its spring snowstorms and convective rainfall. For example, at the Central Sierra Snow Lab near Truckee, California, median April-July (i.e., the most common water supply forecast target period) precipitation over 1991-2020 accounted for only 16.2% of the total annual precipitation; in comparison, at the Copper Mountain SNOTEL in central Colorado, median April-July precipitation across the same period was 31.3% of the total annual precipitation, or nearly double that of California. These two different hydroclimatic regimes, in addition to the geologic differences between the Sierra Nevada mountains of California and the Rocky Mountains of Colorado, highlight some of the challenges with transitioning ASO's technology to a new region and the need for continued study in the state.

As stakeholders consider whether and how the CASM program should expand across the State of Colorado, it is imperative that the leadership team evaluate multiple modeling approaches with and without ASO snow data to determine the performance gains that could be realized with ASO data relative to predefined benchmarks (e.g., historical official water supply forecast skill). This includes working to document experimental streamflow forecasts for ongoing evaluation into the future. To that end, Lynker conducted a retrospective analysis of official streamflow forecasts across the 2022 CASM basins to evaluate long-term (2000-2022) water supply forecast skill (i.e., the reference benchmark for future model inter-comparison studies). We then developed a set of recommendations to guide CASM towards an improved streamflow forecasting program that aims to extract the most information from ASO's high-resolution snowpack data. The results of this first-ever CASM study of seasonal streamflow forecasting in Colorado are summarized herein.

In the western US, seasonal streamflow forecasting, or water supply forecasting, has traditionally been operationally conducted by the National Oceanic and Atmospheric Administration's (NOAA) River Forecast Centers (RFCs) and the U.S. Department of Agriculture's (USDA) Natural Resource Conservation Service (NRCS). Though the two federal agencies employ different techniques and methods for forecasting, they both target seasonal volumes (e.g., April-July) of streamflow at critical forecast points, primarily in snow-dominated regions such as Colorado. While the NRCS operates their Snow Survey and Water Supply Forecasting Program across the western US, NOAA's RFCs are more regionally oriented, with thirteen regional centers across the country. In Colorado, these include the Colorado Basin (CBRFC), the West Gulf (WGRFC), and the Missouri Basin (MBRFC) River Forecast Centers, which forecast for the Colorado River basin, the Rio Grande basin, and the South Platte/Arkansas River basins, respectively.

In support of Northern Colorado Water Conservancy District's (Northern Water) and CASM's Water Plan Grant (WPG) Project Objectives, this study documented and compiled historical official water supply forecasts from the two federal agencies, establishing a proposed framework for model and forecast benchmarks against which future CASM-supported efforts should be measured. This effort is consistent with recent recognition within the hydrologic modeling community that model benchmarking is a critical component for the quantitative and systematic evaluation of new modeling techniques and practices (e.g., Newman et al., 2017), and strives to move beyond the ambiguous stated goals for "improved" or "enhanced" forecasts.

> *"A fully developed ASO program will have accurate, accessible snowpack measurement and improved water supply forecasts across the high-elevation, snow-covered areas of Colorado"*
>
> *-Northern Water's Water Plan Grant Application, Page 7*

This retrospective analysis began at the conclusion of the 2022 Water Year (i.e., September 30th, 2022), following the completion of the 2022 ASO flights and the snowmelt runoff season, which is typically delineated as April 1st to July 31st. The final scope of work, revised at the direction of and with input from the CASM leadership team, focused on the following three areas:

1. A retrospective analysis of 2000-2022 official seasonal water supply forecast products and forecast error, including an empirical error attribution analysis.
2. A hindcast modeling experiment using the SUMMA model (Structure for Unifying Multiple Modeling Alternatives; Clark et al., 2015) from the National Center for Atmospheric Research's Climate & Global Dynamics Lab (NCAR-CGD), leveraging forecasting capabilities developed by Dr. Andrew Wood and collaborators at NCAR-CGD, as a demonstration of an intermediate-complexity hydrologic model.
3. Documentation of spring 2022 experimental forecasts from NCAR's Research Application Lab (NCAR-RAL) WRF-Hydro model and the NOAA Colorado River Basin River Forecast Center (CBRFC) experimental ASO-informed forecasts, for future verification efforts.

Additional deliverables from this task included the facilitation of a "CASM Streamflow Forecasting Roundtable: Water Year 2022" among operational forecasters, modelers, and select forecast end-users (see Sections 8.6 and 8.7 for meeting agenda and minutes) and feedback to ASO Inc. and their experimental forecasting partner, NCAR-RAL, on improvements for better delivery of modeled outputs from the WRF-Hydro model. The outcomes from these conversations are documented in this report along with the 2022 CASM Summary Memo and are actively being integrated into the 2023 CASM program.

3. Methods

3.1. Study Sites

The CASM program, with support from the Colorado Water Conservation Board's (CWCB) Water Plan Grant (WPG), received sufficient grant and match funding to conduct two ASO snow surveys in each of the eight pilot basins across the State in 2022 (Figure 1, blue basins). These pilot basins represented critical locations for key stakeholders in CASM and included the Windy Gap domain (the watersheds of Lake Granby, Willow Creek, and the upper Fraser River), Dillon Reservoir inflows, the upper Gunnison River (East River and Taylor Park Reservoir inflows), Dolores River above McPhee Reservoir, and the upper Conejos River. These study basins have been expanded in 2023 to include the northern Front Range, the upper South Platte, and the upper Roaring Fork basin.

Figure 1: CASM pilot basins (blue) from the 2022 program year, as funded by the Colorado Water Conservation Board's Water Plan Grant and matching funds from stakeholders. Each pilot basin was flown by ASO two times between April-May of 2022, with experimental, ASO-informed streamflow forecasts issued by CBRFC and NCAR-RAL for the 2022 snowmelt runoff season.

From these 2022 pilot basins, Lynker identified and evaluated a total of eight water supply forecast points (Table 1). These study sites were locations that were included in the 2022 ASO snow-on surveys and are existing forecast points for official and/or experimental forecasts from the Colorado Basin River Forecast Center (CBRFC), the U.S. Department of Agriculture's (USDA) Natural Resource Conservation Service (NRCS), and the National Center for Atmospheric Research's Research Applications Lab (NCAR-RAL) WRF-Hydro model. The Conejos River near Mogote, Colorado, as part of the West Gulf River Forecast Center (WGRFC) domain, and the Fraser River at Winter Park, which is not an NRCS forecast point, were the two exceptions. The existing SUMMA models developed by collaborator Dr. Andrew Wood at NCAR's Climate & Global Dynamics Lab (NCAR-CGD) and leveraged in this study were run retrospectively in the East River at Almont, CO and the Taylor River at Taylor Park, CO; no additional SUMMA modeling was conducted in the remaining basins.

Table 1: Study sites and available water supply forecasts from official (CBRFC, NRCS) and experimental (CBRFC, NCAR-RAL's WRF-Hydro, and NCAR-CGD'S SUMMA) modeling systems. Adjusted streamflow volumes are pulled from NRCS historical data, with the exception of the Fraser River at Winter Park, which was calculated from gaged streamflow and diversions.

Basin ID	Forecast Point Name	River Basin	CBRFC	NRCS	WRF-Hydro	SUMMA	USGS Gage	Adjusted Volume Equation
1	East River at Almont	East	X	X	X	X	09112500	Observed Streamflow
2	Taylor Park Reservoir	Taylor	X	X	X	X*	09107000	Calculated Inflow from USBR
3	Willow Creek at Willow Creek Reservoir	Colorado	X	X	X	-	09021000	Calculated Inflow from USBR
4	Lake Granby	Colorado	X	X	X	-	09019000	Calculated Inflow from USBR
5	Dillon Reservoir	Blue	X	X	X	-	09050700	Observed Outflow + (Roberts + Hoosier Tunnel + delta reservoir storage)
6	Dolores River at Dolores	Dolores	X	X	X	-	09166500	Observed Streamflow
7	Fraser River at Winter Park	Fraser	X	-	X	-	09024000	Observed Streamflow + (Jim Creek Diversion)
8	Conejos River near Mogote	Conejos	-	X	X	-	08246500	Observed Streamflow + (delta Platoro Reservoir Storage)

** SUMMA hindcasts were issued for the Taylor Park at Taylor Park forecast point, which is above Taylor Park Reservoir and not inclusive of several smaller tributaries.*

3.2. Snow Data Assimilation (DA)

In the mountainous environments of Colorado and other snow-dominated basins, snowpack storage is a critical means for not only providing water supply during the hot summer months, but also for a source of high predictability of seasonal streamflow volumes. The information that snowpack, along with other watershed hydrologic conditions such as soil moisture, provides to forecasters is key for estimating the runoff (including flood risk) during the spring and summer months. As a result, it is very desirable for forecasters to have an accurate representation of snowpack within the models they use. Data assimilation (DA) of improved snow estimates, such as from ASO, is one approach to improve the simulated snowpack within a model.

In 2022, CASM had access to two experimental real-time forecasting systems that used DA to incorporate ASO's estimates of snowpack into their models. These experimental forecasts, from CBRFC and NCAR-RAL's WRF-Hydro, used a simple DA method called direct insertion to update their simulated snowpack. Direct-insertion is a technique that has long been used by snow modelers (e.g., Essery, 2013 and Hendrick et al., 2018) to take an earth observation (e.g., remotely-sensed snow covered area) or a well-constrained estimate (e.g., ASO snowpack estimates) to "update" the simulated snowpack within a model at a discrete place and time. These updates are performed on the "open-loop" model simulation, which is the model run prior to any DA. The DA model is then run forward in time with these updated model states. While direct insertion is favored for its simplicity, it also poses known challenges. For example, large discrepancies between the simulated snowpack and ASO estimated snowpack during DA, which can occur from both precipitation forcing data errors and lack of model recalibration for optimized snowpack simulation, can result in problematic additions or removals of water from the model. This can lead to unresolved mass balance errors and/or inconsistencies between modeled variables (Magnusson et al., 2016) that translate into inconsistent changes in the forecast through time (e.g., Figure 31, WRF-Hydro forecast). These performance trade-offs are balanced by its computationally efficiency and ease of application. The use of more advance techniques, such as a particle filter, for the assimilation of snow data into hydrologic models has previously shown to provide high value for discharge simulations (e.g., Thirel et al., 2013). These improved DA techniques represent promising opportunities for future CASM research to consider, though they were outside of the scope of work in 2022.

3.3. Historical Water Supply Forecast Evaluation

Water supply forecasts are highly dimensional, often with frequent publication dates (e.g., January 1st, February 1st, etc.), varying target periods (e.g., April-July, April-September), a range of forecast probabilities (e.g., 10-90% exceedance probabilities), and forecasting agencies (e.g., NOAA RFCs and USDA NRCS) even at a single forecast point. For the purposes of this study, and to be most consistent with the experimental forecasts from NCAR-RAL's WRF-Hydro and CBRFC, we evaluated April 1st forecasts targeting the period April 1st - July 31, 2022. Forecasts were compared to NRCS's adjusted streamflow volumes, as documented in Table 1. Our statistical analyses were primarily deterministic, focusing only on the 50% exceedance values (median); where available, we did download and present the full 10-90% probability exceedances for all eight study sites across water years 2000-2022. Exceptions to this included the Fraser River at Winter Park, where no NRCS forecasts have historically been issued, and the Conejos River at Mogote, which is outside of the CBRFC domain. Since WGRFC has only been issuing ESP-based water supply forecasts in the upper Rio Grande for a short period of time, their results are mostly excluded from this study.

In addition to establishing model benchmarks through the historical water supply forecast skill evaluation, we also conducted an empirical analysis of forecast errors. These analyses, presented in Section 4.4, compare forecast percent error as a function of the spring precipitation anomaly, a forecast verification approach used by both the NRCS and the RFCs to evaluate and better constrain sources of forecast error. Spring precipitation anomalies were calculated from the retrospective 1/16th degree GMET forcing dataset (Newman et al., 2015), remapped to a Hydrologic Unit Code (HUC)-12 sub-basin resolution. Due to the limited scope of this study, additional sources of model error (i.e., from diversions, temperature anomalies, etc.) and methods for more directly evaluating forecast error (e.g., model-based sensitivity analysis) were not evaluated here.

3.4. SUMMA Model and Hindcasting Approach

The Structure for Unifying Multiple Modeling Alternatives, or SUMMA, is an intermediate-complexity hydrologic model that was developed by researchers at NCAR in 2015 (Clark et al., 2015) designed to offer a flexible and computational efficient framework for a wide range of use cases. Recently, SUMMA has been successfully applied by researchers at NCAR's Climate & Global Dynamics Lab (NCAR-CGD) to a host of water resource management-specific applications across the western US (e.g., Wood et al., 2021a), including for streamflow forecasting in Colorado (e.g., Wood et al., 2021b) and ASO snow data assimilation in California (Bearup et al., 2021). Recent SUMMA development efforts led by Dr. Andrew Wood at NCAR-CGD and supported by significant and ongoing investment from federal water management agencies including the Bureau of Reclamation and the U.S. Army Corps of Engineers have included:

- Western US-wide 1/16th degree ensemble meteorological model forcings (1970 - yesterday), with limited high-resolution (2 km) implementation, including in Colorado
- Calibration routine; ensemble model hindcasting workflows using ensemble streamflow prediction (ESP) methods
- Model discretization to sub-basin scales optimized for snow process representation (e.g., Figure 2)
- Data assimilation workflows (direct insertion and particle filter), including for ASO snow data
- Real-time ensemble streamflow forecasting system (currently implemented across ~10,000 basins in the Columbia River basin)

For this study, we worked with Dr. Andrew Wood to evaluate the use of the SUMMA model as a demonstration of an intermediate-complexity model for use in CASM-supported hindcasting experiments. Leveraging previous work supported by the Bureau of Reclamation funded projects, we applied calibrated parameters for the HUC-12 scale (i.e., ~100 km^2, or basin-scale; see black outlines of the GRUs in Figure 2) model and supporting modeling workflows to run a series of hindcast experiments in two basins in the upper Gunnison River: the East River at

Almont and the Taylor River at Taylor Park. These demonstrations do not aim to apply the latest SUMMA development work, but rather to demonstrate baseline capabilities of SUMMA and similar intermediate-complexity hydrologic models for providing accurate simulations in a multi-decadal hindcasting framework. Value-added opportunities (including snow data assimilation, model discretization, and high-resolution ensemble model forcings) are beyond the scope of this project and represent future research opportunities for CASM at the discretion of the scientists/researchers who developed these capabilities at NCAR-CGD.

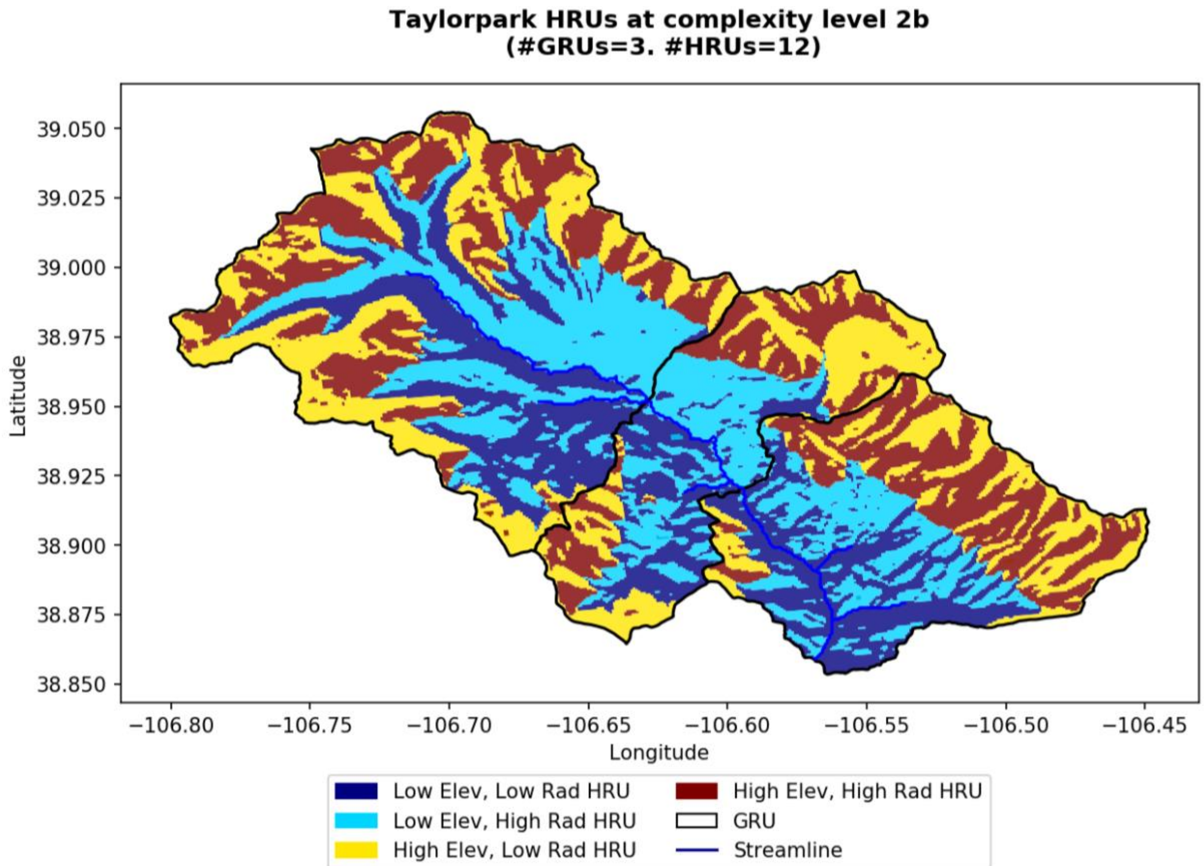


Figure 2: SUMMA model of the Taylor River at Taylor Park, Colorado from Dr. Andrew Wood of NCAR-CGD. The SUMMA model shown here is discretized by elevation bands ('Elev'; high and low) and radiation inputs ('Rad'; high and low) into four hydrologic response units (HRUs; colored) for each geographical response units (GRUs; black outline) for a total of 12 HRUs. This model discretization aims to improve snow process simulations using a computationally efficient strategy; it was not implemented here due to the limited scope of this pilot study.

For our SUMMA hindcast experiments, we first ran the model retrospectively with a single ensemble-member implementation from 1970-2022 to generate a series of model state files (which represent the initial hydrologic conditions of the basin) for the first of every month. Model state files, including information such as snowpack and soil moisture, were then used to re-start, or initialize, the model. For the purposes of this water supply forecasting study, this was done for every April 1st date from 1970-2022. From each of these April 1st dates, we then ran a series of ensemble streamflow prediction (ESP) hindcasts forward in time by one-year (April 1st to March 31st of the following year) using the meteorology from each trace within our dataset (1970-2022), excluding the year of prediction (since that would not be known at the time of a forecast). We then used the MizuRoute model (Mizukami, 2016) to route the simulated runoff from SUMMA. Results were bias-corrected and aggregated for traditional probabilistic spreads (e.g., 10th, 30th, 50th, 70th, and 90th percentiles), volumetric flows (e.g., April – July volumes, in acre-feet), and other metrics of interest (e.g., peak flow magnitude and timing).

4. Results

Empirical results from our official water supply forecast analysis and modeled results from our SUMMA hindcast experiments are summarized in the following section, with supplemental figures and tables included in the appendix. We also document and present the spring 2022 experimental streamflow forecasts that use ASO data from CBRFC and NCAR-RAL's WRF-Hydro for future evaluation efforts pending additional data collection.

4.1. Hydroclimate Data

Table 2 shows basin-scale hydrometeorological trends in the 2022 CASM study basins as measured by a single ensemble-member of the retrospective forcing dataset GMET, or the Gridded Meteorological Ensemble Tool (Newman et al., 2015). For temperature and precipitation variables, we show GMET outputs on both a Hydrologic Unit Code (HUC)-12 scale (i.e., ~100 km^2, or sub-basin-scale) and watershed mean basis; snowpack data (snow water equivalent, or SWE) are pulled from station-based data at nearby NRCS SNOTEL sites. In the East River basin above Almont between the years 1970-2022, for example, we observe strong warming trends at a rate of approximately 0.4° C/decade (0.72° F/decade) and a more moderate decrease in average annual precipitation of -12 mm/decade (-0.47 in/decade), which is also reflected in the Butte SNOTEL measured peak SWE values (-19.4 mm/decade; Figure 3). Similar trends are also observed at other CASM study sites (Table 2; Appendix 8.4), though observations of SWE are often limited by shorter observational records at the nearby SNOTEL stations (e.g., the Upper Taylor SNOTEL 1141 record starts in 2009). Understanding the underlying hydroclimatology is important for contextualizing historical water supply forecast skill, but also for mitigating future skill loss, particularly in basins that have seen dramatic changes in snowpack (e.g., Conejos River).

Table 2: Basin mean hydroclimate statistics from the GMET retrospective forcing dataset (air temperature and precipitation) and the nearest NRCS SNOTEL station with the longest record (peak SWE) for the period 1970-2022 (where data are available).

Basin Hydroclimate		Basin Mean Air Temperature		Basin Mean Precipitation		SNOTEL Peak Snow Water Equivalent		
Basin ID	Forecast Point Name	Average (deg C)	Trend (C/decade)	Average (mm)	Trend (mm/decade)	NRCS SNOTEL	Average (mm)	Trend (mm/decade)
1	East River at Almont	2.55	0.44**	569	-12.3	380	397	-19.4
2	Taylor Park Reservoir	1.30	0.47**	536	-10.0	680	391	-9.35
3	Willow Creek Reservoir	2.30	0.29**	485	-12.2	869	404	10.1
4	Lake Granby	1.80	0.34**	656	2.0	565	701	-33.1
5	Dillon Reservoir	0.74	0.39**	648	-5.6	415	423	8.9
6	Dolores River at Dolores	4.33	0.41**	679	-33.0*	1185	418	-26.4
7	Fraser River at Winter Park	1.09	0.33**	761	-9.7	305	579	-23.5
8	Conejos River near Mogote	2.04	0.53**	836	-40.4*	580	403	-49.9*
Average		2.02	0.40	646	-15.1	-	465	-17.9

**statistically significant trend at p-value 0.05*
***statistically significant trend at p-value 0.01*

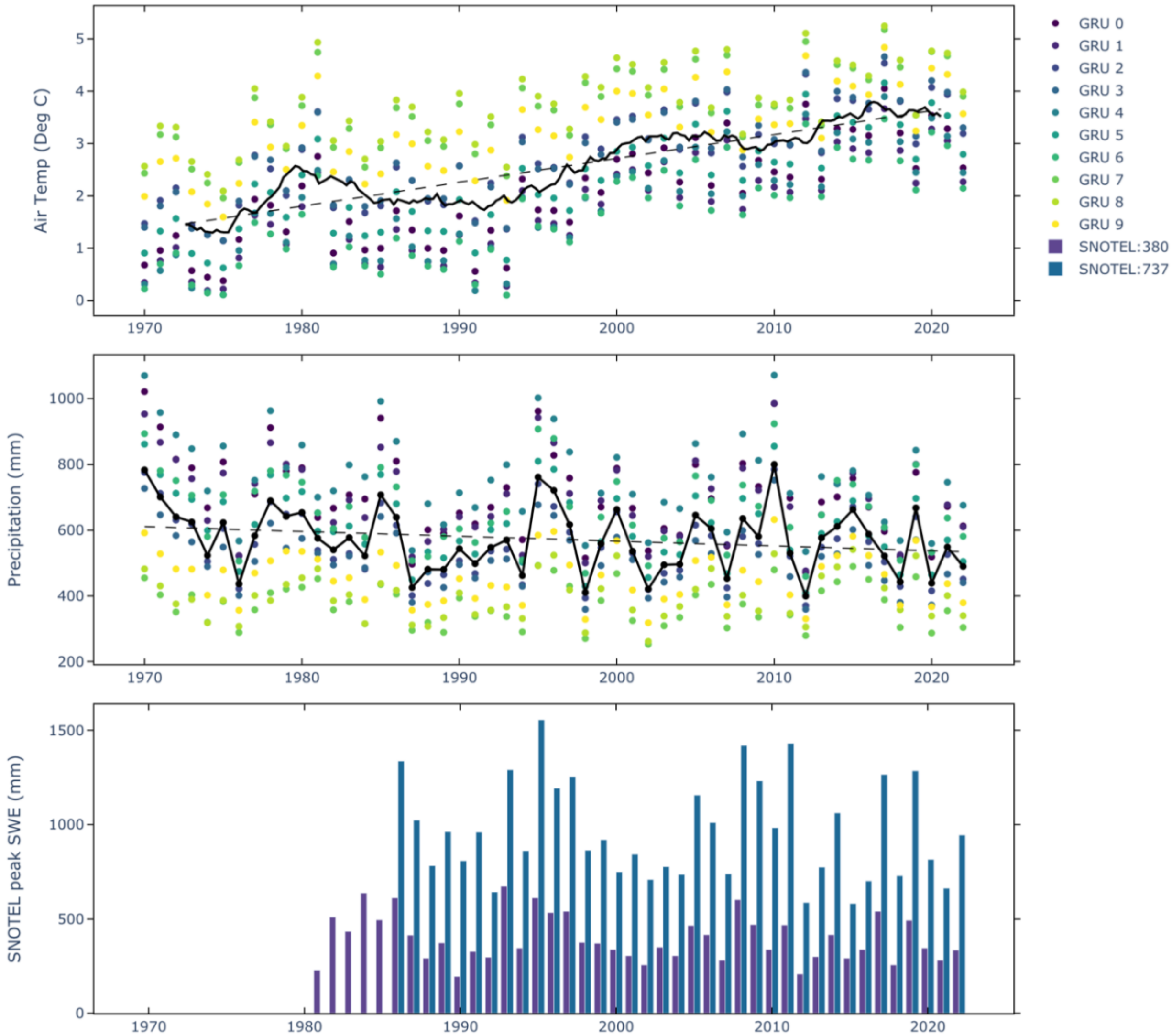


Figure 3: East River at Almont forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL stations. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL 380 and 737; bottom panel). Temperature and precipitation are measured on a sub-basin level (HUC12 basins, labeled as GRUs, or Geographical Regional Units), as shown by the colored dots, with basin mean conditions denoted by the black line. The East River at Almont consists of ten GRUs, or HUC12 basins. Grey dashed lines show air temperature and precipitation trendlines; black solid lines show a five-year rolling average.

4.2. Historical Water Supply Forecast Skill

Improving seasonal water supply forecasts is contingent upon determining first the benchmark against which to systematically evaluate new modeling approaches. To better establish and promote model benchmarking by CASM and its partners, we present summary statistics from official water supply forecasts from both the NRCS and the CBRFC for the eight study basins across the period 2000-2022 (Table 3). These statistics are calculated on the median value of all April 1st forecasts of April-July adjusted volumes, as reported by NRCS. Full probabilistic distributions of the forecasts, which show greater nuance on a basin-year basis, are presented in the boxplots, e.g., the East River at Almont in Figure 4 (NRCS forecasts are on the left panel; CBRFC forecasts are on the right). All other forecast boxplots for the study basins can be found in Section 8.2 in the Appendix. The full probabilistic distributions are important to consider since a forecast is not always expected to verify at the 50% exceedance level. For example, if the April-July period is anomalously hot and dry, we would expect a forecast issued on April 1st to verify closer to the 70 or 90% exceedance level (i.e., the bottom of the forecasted range).

Table 3: April 1st water supply forecast skill from NOAA's RFCs and USDA's NRCS for the period 2000-2022, as measured by the correlation coefficient (r^2), mean absolute error (MAE; in thousand-acre feet; KAF), and percent bias (PBIAS), relative to the 50% exceedance value of the forecasts. Basin average statistics are presented for the normalized error metrics (i.e., r^2 and PBIAS).

Basin Statistics			Correlation Coefficient (r^2)		Mean Absolute Error (MAE; KAF)		Percent Bias	
Basin ID	Forecast Point Name	Average AMJJ Vol. (KAF)	NOAA RFC	NRCS	NOAA RFC	NRCS	NOAA RFC	NRCS
1	East River at Almont	159	0.83	0.86	22	20	1.79	2.10
2	Taylor Park Reservoir	84	0.74	0.78	13	11	3.45	4.02
3	Willow Creek at Willow Creek Reservoir	49	0.70	0.64	12	12	-11.96	-7.06
4	Lake Granby	217	0.72	0.73	32	31	-7.03	-5.22
5	Dillon Reservoir	153	0.80	0.80	21	21	2.51	4.39
6	Dolores River at Dolores	185	0.80	0.79	35	34	8.23	7.69
7	Fraser River at Winter Park	18	0.46	N/A	3	N/A	7.73	N/A
8	Conejos River near Mogote	148	N/A	0.80	N/A	22	N/A	0.80
Average		**126**	**0.72**	**0.77**	**-**	**-**	**0.67**	**0.96**

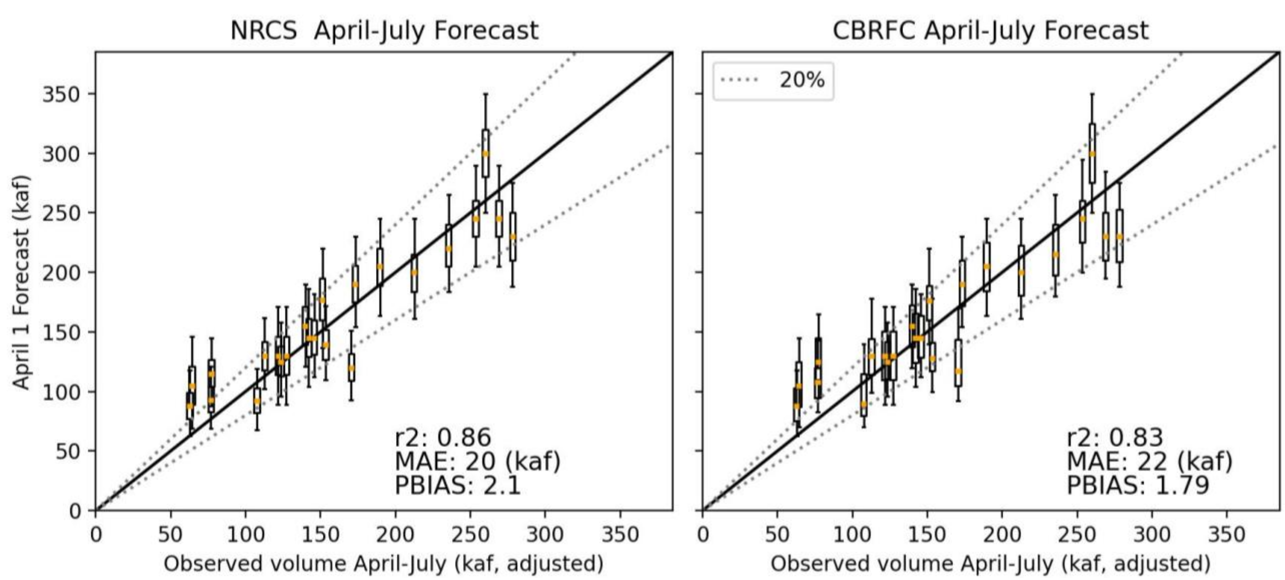


Figure 4: East River at Almont April 1st forecasts of April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right) in units of thousand-acre-feet (KAF). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, orange dot notes the 50% level (median). Dashed lines represent the 20% bounds around the 1:1 line.

4.3. Experimental Forecast Documentation

Forecast verification is an essential practice that is regularly performed by the federal operational forecasting agencies when trying to evaluate and explain their past forecast performance (e.g., [CBRFC water supply verification webpage]). Recognizing the value of such retrospective assessments, a verification analysis of CASM-supported experimental forecasts was an initially scoped aspect of this study. However, after further discussion among members of the CASM planning team, it was determined that with such a limited ASO snow dataset, an extensive forecast verification exercise focusing on a single year would be premature. Accordingly, our efforts instead focused on documenting experimental forecast products that used ASO snow data assimilation (from NCAR-RAL's WRF-Hydro and CBRFC) for future evaluation pending additional years of ASO data.

With future verification efforts in mind, we compiled tabulated and graphical representations of experimental and official forecasts for the eight forecast points within this study for water year 2022 (see tables in Section 8.3 of the Appendix). Because NCAR-RAL's WRF-Hydro forecasts did not regularly provide probabilistic forecast ranges in 2022, we generally focused on tabulating and visualizing comparable forecasts of the 50% exceedance (median) values between all agencies and modeling groups (Table 4). Where available (e.g., from official forecasting entities), we also document the full probabilistic spread of forecasts. Figure 5 shows an example of the East River at Almont water supply forecasts from official (NRCS and CBRFC) and experimental ASO-informed (CBRFC and WRF-Hydro) systems. Figures for all other study basins are in Section 8.3 of the Appendix. It is important to note that in water year 2022, WRF-Hydro's late-April forecasts were issued for a target period of April-September, not April-July (which is the official forecast period for all study sites except for the Conejos River at Mogote). Since individual traces from WRF-Hydro forecasts were not publicly available and not made available in support of this study, we estimated April-July volumes from the provided April-September volumes by subtracting the NRCS-reported average August-September volume (which was generally 10% or less of the total volume for the period); see Table 6 and Table 7 in the appendix for further details. Other complicating factors in comparing spring 2022 experimental forecasts included inconsistent training datasets (e.g., observed flows vs. naturalized flows) and representation of managed systems across models and forecasting systems. For example, WRF-Hydro, unlike CBRFC and NRCS, does not explicitly represent consumptive use, diversions, or reservoirs. This includes the representation of reservoirs such as Willow Creek and Lake Granby as natural lakes as compared to managed reservoirs with operational rules, which are explicitly represented within CBRFC's conceptual models, and implicitly represented within NRCS's statistical models. NCAR-RAL is working to re-calibrate and re-configure the model for better performance in and representation of managed basins in future years. This discrepancy is very evident in the poor model performance relative to observations in highly managed systems, such as Lake Granby (Figure 27).

In general, however, forecasts that are closer to the black horizontal line (which is the NRCS-reported April-July adjusted volume, unless otherwise noted) than the red dash line (which is the NRCS-reported median April-July volume for the period 1991-2020) are an improvement over the climatology. Important caveats include that 1) a single forecast year is generally insufficient to adequately assess true model skill and, 2) a forecast can be considered skillful even if it verifies at the upper or lower end of its probability distribution (as denoted by the small vertical error bars). For these reasons, we refrain from comparing forecast performance, and instead focus on documenting the real-time experimental forecasts from spring 2022 and advocating for future CASM-supported hindcast modeling studies that control for these many limitations.

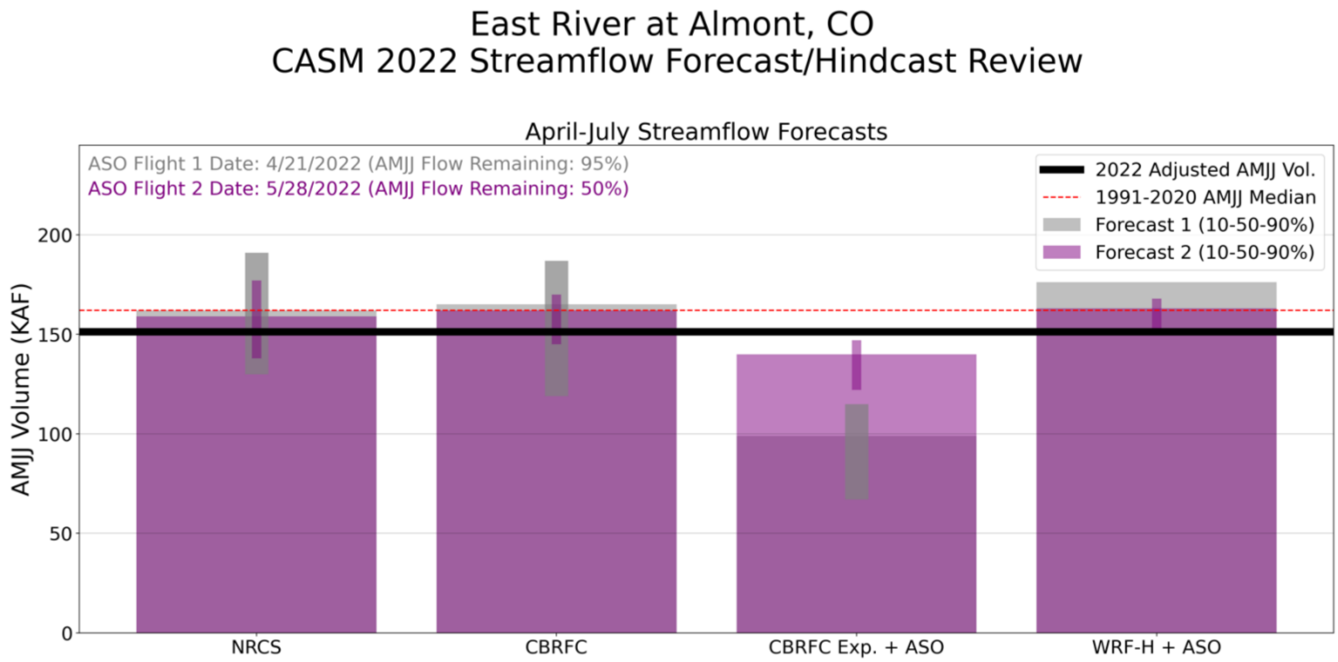


Figure 5: Official and experimental water supply forecasts for the East River at Almont, showing the median April-July forecasted volumes in 2022 (in thousand acre-feet, or KAF) bounded with error bars by the 10-90% forecast exceedance volumes. 'Forecast 1' (grey) and 'Forecast 2' (purple) issue dates all vary slightly across models. For example, depending on the model, 'Forecast 1' corresponds to either a) the first ASO flight date (here, April 21st, 2022), b) the nearest official forecast date relative to the ASO flight time (e.g., May 1st for the NRCS and CBRFC forecasts in grey). The black horizontal line is the NRCS-reported adjusted April-July volume; the red horizontal dashed lines is the NRCS-reported 1991-2020 median volume. Late-April WRF-Hydro forecasts were post-processed from the provided April-September volumes to April-July volumes by subtracting out NRCS-reported average flows for August and September; see Table 6 and Table 7 in the Appendix for additional information.

Table 4: Water supply forecast error (right of the black bar, as a percent of NRCS-reported April-July adjusted volumes) for the 50% exceedance value from the April 1st forecast/hindcast (NRCS, CBRFC) or the nearest first ASO flight (CBRFC, WRF-Hydro; generally in late April). Note that the ASO-informed forecasts from CBRFC and NCAR-RAL were issued towards late April (when ASO was flown), which is nearly a month after the NRCS and CBRFC April 1st forecasts.

Basin ID	Forecast Point Name	2022 Adj. Vol. (KAF)	NRCS	CBRFC	CBRFC + ASO	WRF-Hydro + ASO*
1	East River at Almont	151.1	17.1	16.4	-34.5	16.6
2	Taylor Park Reservoir Inflows	71.9	21.0	33.5	2.9	-2.3
3	Willow Creek at Willow Creek Reservoir	67.0	-31.3	-28.3	-38.8	**
4	Lake Granby Inflows	221.2	-9.6	-11.8	-24.5	**
5	Dillon Reservoir Inflows	112.5	22.7	20.0	14.7	0.4
6	Dolores River at Dolores	137.4	7.0	0.4	27.3	18.1
7	Fraser River at Winter Park	15.6	-	3.7	-6.2	-7.6
8	Conejos River near Mogote	169.7‡	-5.1	-	-	0.9

**NCAR-RAL's WRF-Hydro 2022 forecasts were issued as natural flows (i.e., without management), which in managed systems, is inconsistent with the adjusted volumes published NRCS; these are generally calculated as observed flows, less any diversions, or as reservoir outflows plus changes in the reservoir storage. Heavily regulated systems include Lake Granby, Dillon Reservoir, and the Conejos River at Mogote, all of which have upstream reservoirs that are explicitly (CBRFC) or implicitly (NRCS) represented by official forecasting agencies. WRF-Hydro was updated to represent lakes and reservoirs in 2023. Furthermore, WRF-Hydro ASO flight one forecasts (late April) were issued for the period April-September and had to be post-processed using climatology to estimate April-July forecasts. The August-September flows were generally small at around 10% or less of the total forecasted volume for the April-September period.*

***WRF-Hydro forecasts are not available due to technical issues in the NCAR-RAL forecasting system. Additionally, in 2022, Willow Creek Reservoir and Lake Granby were represented as natural lakes, not reservoirs, which is reflected in the poor performance of May WRF-Hydro forecasts (e.g., Figure 27).*

‡Due to the monsoon season, the Conejos at Mogote is traditionally forecasted for April-September, which is what is reported here.

4.4. Forecast Error Attribution: Spring Precipitation Anomalies

Spring precipitation anomalies are a well-recognized source of water supply forecast error, particularly for early-season forecasts with variable precipitation patterns. In our communications with forecasters from NRCS and CBRFC, it was evident that these forecasting agencies regularly evaluate the performance of their water supply forecasts with respect to anomalous hydroclimatic conditions, including precipitation and temperature, to verify that the forecasts performed as expected. For example, if April-July precipitation is far below the long-term average of the historical training data used to calibrate the model, then forecasters would generally expect the 50% exceedance forecast (i.e., the median) to over-forecast, through the hydrologic conditions of the basin (including, e.g., soil moisture deficits) will modulate the hydrologic response to precipitation.

Our empirical analyses of official April 1st water supply forecasts from the NRCS and CBRFC for water years 2000-2022 showed that forecast error (as measured as a percent of the observed volume) was generally well-correlated with GMET-measured precipitation anomalies during the forecast target period, April 1st-July 31st. On average, the Pearson r^2 between these two variables was approximately 0.5 across both forecasting agencies (Table 5); in other words, 50% of the forecast errors could be statistically described by April-July precipitation anomalies. Individual forecast errors by basin ranged from highly-correlated in the Lake Granby basin (upwards of 0.7; Figure 35) to moderately well correlated in the East River (about 0.5; Figure 6) to poorly correlated in the Conejos River (0.3; this low r^2 value was driven in part by a single outlier during an exceedingly dry year; Figure 39). These results are largely consistent with a CBRFC model-based sensitivity analysis of four of their forecasts from April 1st, 2022 showing that approximately half of their forecast error was from spring precipitation (Figure 7). Empirically evaluating historical water supply forecast error provides valuable information about where error might be coming from, how we might better address it in our models, and what the ceiling for forecast improvement is when employing methods that don't constrain precipitation anomalies, e.g., ASO snow data assimilation. Additional spring precipitation forecast error analyses can be found in Section 8.4 in the appendix.

Table 5: Spring (April-July) precipitation anomalies as measured by the retrospective GMET forcing dataset and correlation coefficients with forecast percent errors for both NRCS and CBRFC April 1st forecasts for the period 2000-2022.

Basin Statistics		Spring (AMJJ) Precipitation Anomaly (mm)		Explanatory Power of Spring Precipitation Anomaly (r^2)	
Basin ID	Forecast Point Name	Minimum	Maximum	NOAA RFC	NRCS
1	East River at Almont	-85	110	0.58	0.60
2	Taylor Park Reservoir	-83	129	0.56	0.47
3	Willow Creek at Willow Creek Reservoir	-90	166	0.49	0.50
4	Lake Granby	-114	231	0.68	0.71
5	Dillon Reservoir	-99	218	0.53	0.50
6	Dolores River at Dolores	-90	146	0.50	0.50
7	Fraser River at Winter Park	-136	200	0.32	-
8	Conejos River near Mogote	-115	185	-	0.30
Average		**-101**	**173**	**0.52**	**0.51**

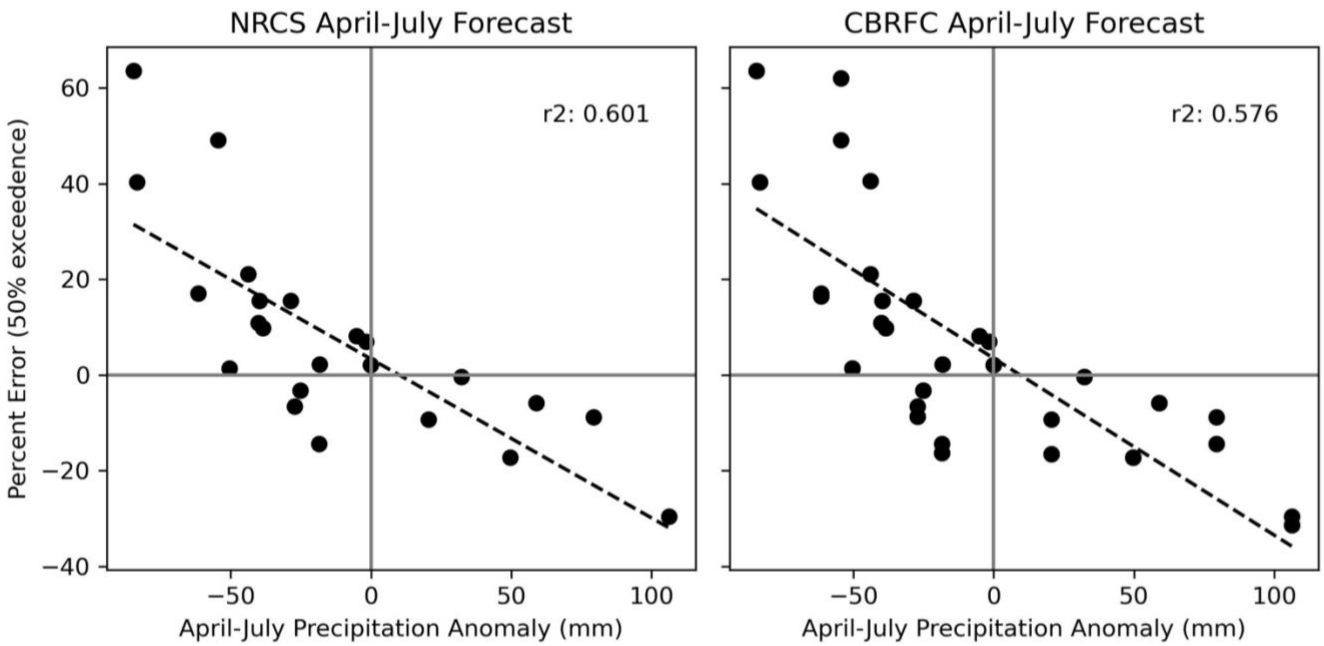


Figure 6: April 1st forecast percent error (for the 50% exceedance forecast of April-July adjusted volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right) for the East River at Almont (2000-2022), as measured by a single-ensemble member of the GMET retrospective forcing dataset. Results show that about 60% of forecast error in the East River is explained by spring precipitation anomalies.

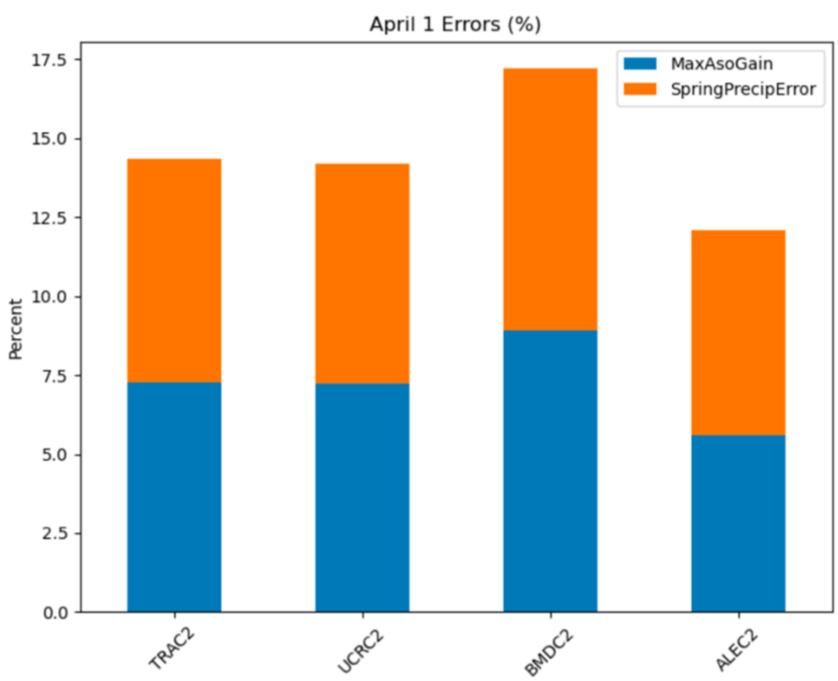


Figure 7: A figure from CBRFC's Water Year in Review report showing CBRFC-estimated spring precipitation errors (orange) for April 1st forecasts of April-July volumes across four basins in Colorado (NOAA, 2022). TRAC2: Taylor Park, UCRC2: Uncompaghre near Ridgeway, BMDC2: Blue Mesa Reservoir, ALEC2: East River at Almont. Results show that approximately 50% of forecast error is from unknown spring precipitation (orange) at the time of the forecast. Blue bars show error from other sources, such as simulated snowpack and/or model structural errors.

4.5. SUMMA Hindcast Experiment

SUMMA modeling was conducted across two upper Gunnison River headwater basins as a part of a pilot study for evaluating the suitability of other hydrologic models for the CASM streamflow forecasting program. Since SUMMA was not included as part of the CASM 2022 real-time experimental streamflow forecasting season, all SUMMA simulations were conducted retrospectively. These hindcasts leveraged development work led by researchers at NCAR-CGD, namely, Dr. Andrew Wood. Hindcasting, or "re-forecasting", is a commonly used scientific approach that provides an opportunity to retroactively study the performance of a model as if it had been run as a real-time forecast initiated on, for example, April 1st, 2022. Additionally, hindcasting allows us to issue re-forecasts in a controlled way across multi-decadal periods, which is critical for studying long-term model performance, assessing incremental improvements, and for evaluating and correcting for systematic biases.

In partnership with researchers at NCAR-CGD, SUMMA model results demonstrate baseline capabilities of intermediate-complexity/scale hydrologic models for providing high-quality simulations in a single year & multi-decadal hindcasting framework.

At both the East River at Almont and the Taylor River at Taylor Park, 2022 SUMMA model results indicated good performance across both the ensemble streamflow prediction (ESP) hindcast and the retrospective model runs (Figure 8 and Section 8.5), with Nash Sutcliffe-Efficiency values of greater than 0.8 for bias-corrected ESP hindcasts (Table 10 and Table 11). Importantly, the ESP hindcasts (e.g., Figure 8, light grey traces, with shaded red 10-90% probability distributions from years 2000-2021) demonstrate performance of the hindcast without the 2022 meteorology, since this was an unknown on April 1st, 2022. In contrast, the retrospective model simulation (blue line) shows actual model performance if we had known the hydroclimate for the forecast target period. Comparison of the median ESP hindcast (red line) with the retrospective simulation (blue line) provide the difference in model performance with and without the actual observed weather for 2022 (i.e., the equivalent of the 'SpringPrecipError' in orange; Figure 7); remaining error in the retrospective model simulation (i.e., the equivalent of the 'MaxAsoGain' error in blue; Figure 7) show performance improvement opportunities, including across model structure (e.g., better calibration), model inputs (e.g., forcing data), and model initial conditions (e.g., simulated snowpack or soil moisture). Additional model enhancement via snowpack data assimilation, use of high-resolution (e.g., 500 meter) model forcings, and spatial discretization of the model for optimized representation of snowpack processes were outside of the scope of this project, but are existing demonstrated capabilities of SUMMA.

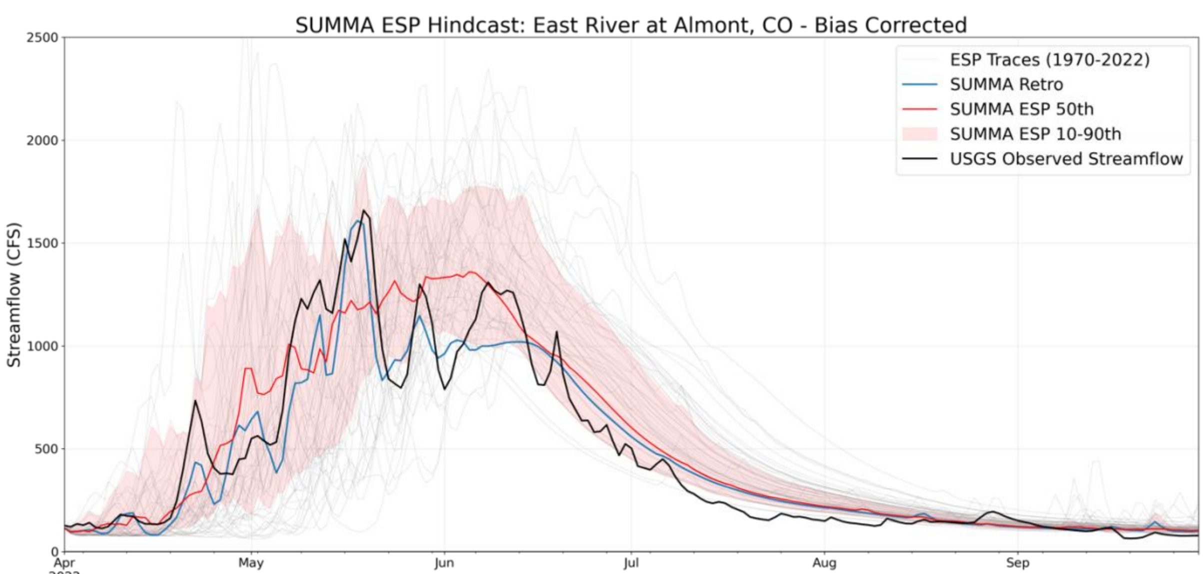


Figure 8: SUMMA bias-corrected ESP hindcast of the East River at Almont for April 1st, 2022, showing 53 ensemble members from meteorological years 1970-2022 (grey traces) with respect to observed flows from the USGS gage (black). Probability distributions for meteorological years 2000-2021 are highlighted in red (10-90th percentile) where the median (50th percentile) is denoted by the solid red line. The 2022 retrospective model run (which uses actual 2022 forcings) is in blue.

While evaluating model performance across a single year is informative, it is also important to evaluate longer-term (e.g., multi-decadal) trends in hindcast skill, an advantage of more computationally efficient intermediate-complexity hydrologic models which can be run using many ensemble members across long simulation periods. In Figure 9, we show a concurrent series of 22 one-year ESP hindcasts initiated on April 1st, ranging from April 1st, 2000 to April 1st, 2022; forecast issue dates are marked by the orange circles. From these 22 hindcasts, we then aggregate bias-corrected volumetric April-July flows: the 10th – 90th percentile exceedances in the boxplot on the bottom panel, where the deterministic retrospective simulations (blue line in Figure 8) are denoted by teal squares. Relative to historical April 1st water supply forecast skill from CBRFC and NRCS across the same period (Table 3 and Figure 4), SUMMA ESP hindcasts of April-July volumes demonstrate comparable skill, with similar or better r^2, percent bias (PBIAS), and mean absolute error (MAE) values. These results suggest the good suitability of SUMMA and similar models for continued study as part of CASM, including ASO snow data assimilation and improved snow process representation through discretization (e.g., Figure 2), both of which were outside of the project scope of work.

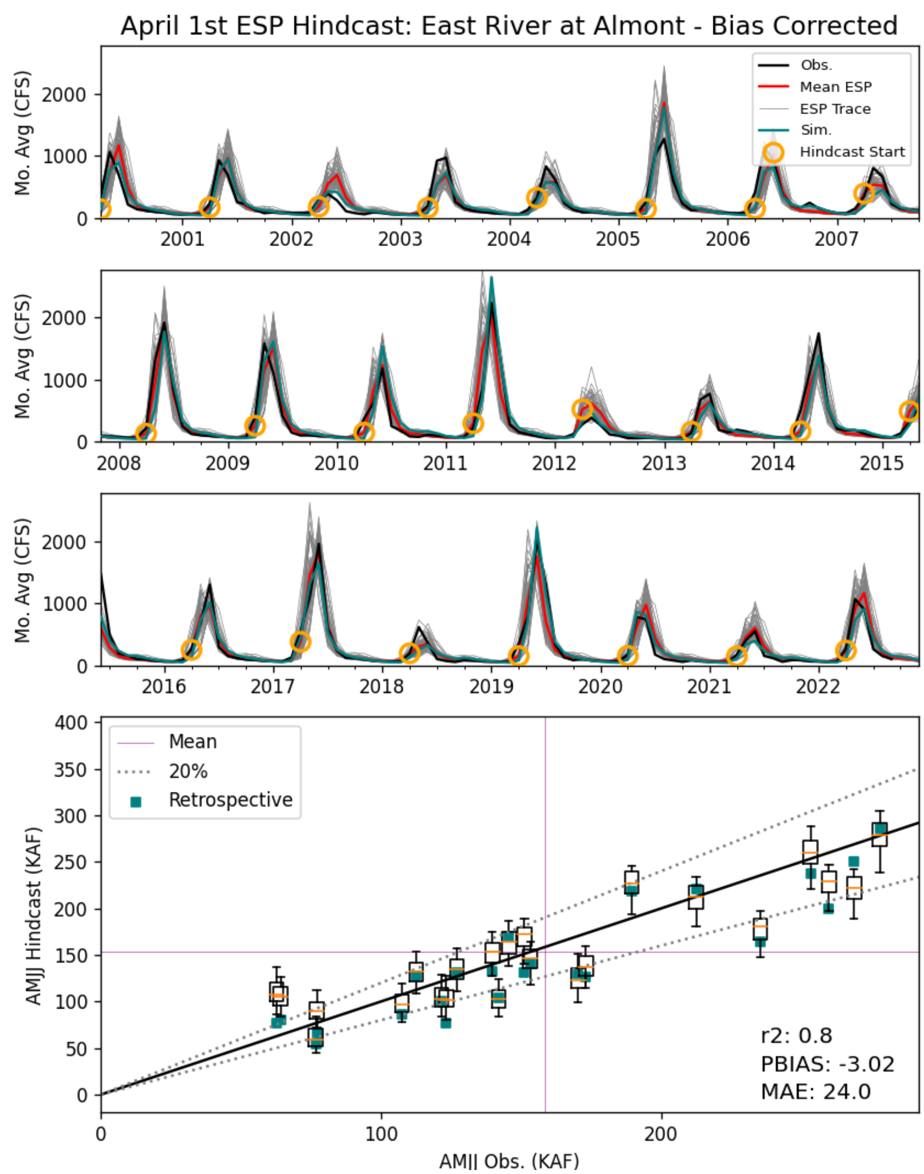


Figure 9: Multi-decadal SUMMA hindcasts (leave-one out bias corrected-ESP for all April 1st dates) and retrospective simulations for the East River at Almont. Here, the top panels show the 2000-2022 time series of the one-year April 1st hindcasts; the bottom panel shows the volumetric April-July flows for the probabilistic hindcasts (boxplots) and deterministic retrospective simulation (teal square) against USGS streamflow observations. Statistics are of the median BC-ESP hindcast.

5. Discussion

Past evaluations of western US water supply forecast (WSF) skill have demonstrated that in the Colorado River basin, operational forecast skill from statistical methods has been declining since the 1980s (Pagano et al., 2004). Research has shown that this period has been characterized by an increase in both streamflow variability and persistence (e.g., extended dry and wet periods; Pagano and Garen, 2005) and changes in streamflow timing (Stewart et al., 2005) and magnitude (Lins and Slack, 1999). These hydroclimatic trends have negative implications for WSF and water management in the western US more broadly and are only further compounded by the related forces of increasing temperatures (Lehner et al, 2017) and declining snowpack (Livneh and Badger, 2020; Harpold et al., 2017). Improving WSFs and/or developing more climate change resilient forecasts has been a hot topic of academic and operational research of late. This has included evaluating the use of climate information such as seasonal temperature forecasts (Lehner et al., 2017) and climate indices (Mendoza et al., 2017) in existing model frameworks, the assimilation of high-resolution ASO snowpack data into hydrologic models such as LIS/WRF-Hydro (Lahmers et al., 2022) and SUMMA (Bearup et al., 2021), and through the use of new modeling techniques altogether, such as machine learning, by both federal agencies (Fleming et al, 2021) and commercial companies such as Upstream Tech and Google's Flood Forecast (Nevo et al., 2022).

The research-to-operations gap is so vast in the field of hydrologic prediction that NOAA recently established a $360M cooperative institute (CIROH) committed to advancing water prediction. CASM should aim to collaborate with such groups and people.

Motivated by technological advancements and observed climatic change, there has been recent stakeholder-driven interest in efforts to improve streamflow forecasting across nearly all forecast time horizons, but especially for seasonal WSF. Numerous efforts from the state/local level, such as CASM in Colorado, to the national level, such as those by the Bureau of Reclamation, are currently underway to develop a more climate resilient and improved WSF. Conversations with several operational forecasters as part of this study has revealed that historically, these efforts have underestimated the difficulty of improving upon their operations. To help address this well-known research-to-operations gap, NOAA recently established a $360 million institute tasked with addressing these challenges specifically: the Cooperative Institute of Research to Operations in Hydrology (CIROH, 2023). CIROH recognizes that transitioning research workflows to operational use requires not just skillful model guidance (a well-recognized challenge for the WRF-Hydro-based National Water Model, NWM), but also the maturity/stability of the model framework (which is currently being addressed through the development of the NextGen NWM at NOAA's National Water Center, NWC). CIROH aims to facilitate this research-to-operations, or "R2O", transition by advancing community water modeling and collaborative research, including research on the use of the NextGen NWM by groups outside of the NWC. CASM should consider how its efforts to improve hydrologic prediction through ASO snow data assimilation fits into this unprecedented effort in the hydrologic sciences.

To improve upon existing methodologies in Colorado, whether by using ASO snowpack data assimilation or other approaches, will require a systematic approach by CASM and its partners that works to 1) understand and evaluate the existing skill of operational WSFs in Colorado, 2) acknowledge and adapt/apply the extensive past and current research of government, academia, and industry in the field of hydrologic prediction, and 3) employ the scientific method to measure incremental improvements over established benchmarks, ideally in a controlled model hindcasting framework and through peer-reviewed channels.

In support of these goals, the first objective of this study was to document and assess historical WSF skill from the primary operational forecasting agencies for the CASM study basins: NOAA's Colorado Basin River Forecast Center (CBRFC) and the USDA's Natural Resource Conservation Service (NRCS). Our results showed that across our study basins, average April 1st forecast error between the years 2000-2022 for the 50% exceedance forecast was approximately 16% for April-July volumetric flows. Lynker's analysis of historical WSFs showed that these errors were well-correlated with spring precipitation anomalies for the April-July months: intuitively, forecasts systematically over-predicted streamflow during dry spring periods, and systematically under-predicted during wet spring periods. While the explanatory power of April-July precipitation anomalies varied across our study domain, our results generally showed that 50-70% of WSF error could be explained (empirically) by this variable alone, with exceptions in the Fraser River and Conejos River basins, where explanatory power was lower (the period of record in the Fraser is much shorter, while a dry outlier year in the Conejos weakens the relationship). These findings were consistent with results from CBRFC which estimated that for the four basins evaluated in their sensitivity analysis, 50% of their forecast error in 2022 was from spring precipitation anomalies (NOAA, 2022). The operational forecasting community widely recognizes post-forecast precipitation anomalies as a large and mostly unresolved source of error (though seasonal to sub-seasonal, or S2S, climate models are making gains here, as are shorter-range numerical weather predictions). Nonetheless, it highlights one of the key challenges with transitioning ASO's technology and streamflow forecasting approach from the Mediterranean climate of California (where spring precipitation is minimal) to Colorado (where a larger fraction of the total annual precipitation falls during the spring runoff season). Better evaluating these sources of existing forecast error, and what they mean in terms of the challenges and opportunities for improving our simulations of one of the most critical components of Colorado's water balance — snowpack — on both an annual and long-term basis should be paramount for CASM and its stakeholders.

These types of empirical analyses are critical for not just understanding sources of forecast error, but also opportunities for improving WSF skill. Hindcast modeling, in which modelers issue "re-forecasts" after the fact, provides yet another tool for answering these types of research questions. For example, in the modified figure from Menodza et al., (2017) (Figure 10), we can see the comparative value of watershed information (e.g., snowpack and soil moisture conditions) with climate information (e.g., climate indices, including El Nino Southern Oscillation, ENSO) across forecast lead time when predicting spring runoff into Dworshak Reservoir, Idaho. Using hindcasting to compare thirteen different modeling approaches, Mendoza et al. (2017) demonstrate the skill of different techniques that use only watershed information (black and grey bars), only climate information (orange bars), and a hybrid approach of the two (red, purple, blue, and green bars). At the beginning of the water year, when little to no snow has accumulated, climate information provides the most predictability for April-July runoff. However, as the water year progresses and the snow accumulates, and we near April 1st, we see that methods relying on watershed information alone (e.g., ensemble streamflow prediction, ESP, in black) provide more skill relative to climate information alone, with the most robust forecasts coming from methods that use all available information (e.g., the ESP trace weighting scheme, TWS, in blue). Similar hindcast experiments that control for the use of ASO snow data assimilation in WSFs would be enormously beneficial to CASM and their stakeholders and improve programmatic credibility.

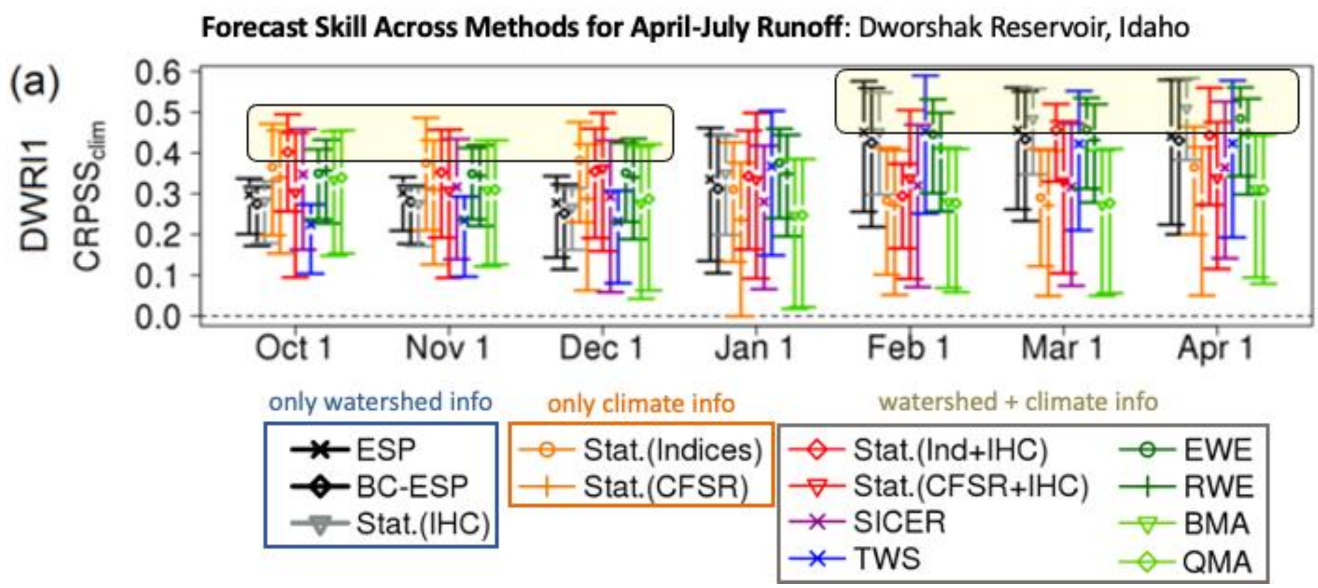


Figure 10: Modified figure from Mendoza et al. (2017) comparing water supply forecast skill (as measured by the Continuous Ranked Probability Skill Score, CRPSS, where higher scores are better) for hindcasts of April-July runoff into Dworshak Reservoir, Idaho. Results show that early in the water year, we can improve forecasts using climate information, while closer to April 1st, we can develop the most robust forecasts by using both watershed (e.g., snowpack) and climate (e.g., El Nino Southern Oscillation indices) information as shown by the ensemble streamflow prediction trace weighting scheme (TWS) in blue. CASM should aim to replicate similarly rigorous scientific modeling studies controlling for the use of ASO snow data in water supply forecasts.

We used NCAR-CGD's SUMMA model to run a series of hindcast experiments in two headwater basins of the upper Gunnison River to perform our own selective hindcast experiments that aimed to show the utility of retrospective studies with intermediate-complexity and scale hydrology models. As compared to the high-resolution (e.g., one-kilometer spatial resolution) and high-process representation of WRF-Hydro and similar models, such as DHSVM, intermediate-complexity models run at a coarser spatial resolution (e.g., VIC at 1/8th to 1/16th degree) have the advantage of being more computationally efficient and nimble. These intermediate-complexity and scale models still retain greater process realism over the highly calibratable conceptual hydrologic models, such as Snow17/SAC-SMA (used by NOAA's River Forecast Centers) or SWAT in Europe (e.g., Abbaspour et al., 2015) on the other end of the modeling spectrum, though less than models such as WRF-Hydro or DHSVM. This balance of physical process representation and computational efficiency allows for the implementation of and experimentation with important forecasting and hindcasting capabilities, including multi-decadal ESP model runs (e.g., Figure 9) and improved data assimilation techniques (such as a particle filter), all within a physically-oriented hydrologic modeling framework that retains enough model fidelity for critical hydrologic process representation.

Though geographically constrained, our pilot study demonstrated the high suitability of intermediate-complexity models such as SUMMA for ensemble seasonal streamflow forecasting in Colorado. In addition to a targeted hindcast experiment for water year 2022 (Figure 8), we also demonstrated the ability of intermediate-complexity models to be used in a multi-decadal hindcasting studies (Figure 9), which is critical for establishing metrics for long-term model performance, but also for applying statistical post-processing corrections (e.g., bias correction). This hindcast experiment was the first-of-its-kind by a CASM-supported project. Despite minimal model calibration (due to limited project scope) and our decision to apply an intermediate-resolution sub-basin (HUC12) scale of the model not yet optimized for snow process representation and without snow data assimilation, the bias-corrected SUMMA hindcasts for both the East and Taylor Rivers were, by all measures, quite skillful: for the 22 hindcast years of this study (2000-2022), correlation coefficient values (r^2) of April-July volumes were ~0.8, with an absolute bias of less than 3% and a mean absolute error similar to or less than real-time forecasts from CBRFC and NRCS. Increased model spatial resolution through model discretization to hydrologically similar response units (e.g., Figure 2) run with high-resolution model forcings and assimilation of ASO snowpack data (all of which are existing capabilities that were outside of the scope of this study) would likely improve upon these performance measures.

These results reinforce the value of using process-oriented hydrologic models run in ensemble hindcast modes for these types of studies. While our results are only precursors, they highlight the need for additional modeling studies by CASM and its partners, ideally published in peer-reviewed scientific journals, with an experimental design optimized for evaluating programmatically imperative research questions including: what is the marginal benefit of high-resolution snowpack data in seasonal WSFs? Are there better ways to derive greater value from ASO snowpack data and other earth observations into hydrologic models beyond direct insertion? Why might the direct insertion of improved snowpack data not yield improved hydrologic predictions of streamflow, given the undeniable accuracy of lidar-based snowpack measurements? And can we calibrate our way to better snow data assimilation for improved streamflow predictions? Other research questions that CASM might consider, which extend beyond ASO snow data assimilation, could include: How might climate information such as sub-seasonal climate forecasts be used in operational or experimental models for addressing sources of model error related to spring precipitation (or temperature) anomalies? Or, how do LSTM-based machine learning models compare to processed-based hydrologic models? All of these questions are deserving of exploration before statewide expansion of the CASM program.

6. Conclusions and Recommendations

It is evident that challenges remain in the effort to translate improved snowpack estimates from ASO into reliably enhanced water supply forecasts for the State of Colorado, though numerous opportunities and research tracts remain. On the path toward this goal, CASM should focus on evidence-based approaches that allow for direct comparison of experimental approaches with WSFs, or other key benchmarks. Despite mixed evidence for the use of directly-inserted ASO snow data in streamflow forecasting models for Colorado in 2022 (NOAA, 2022 and Section 8.3 of this report), data assimilation of high-resolution snowpack information may well provide one avenue, among others, to help achieve this goal of WSF improvement. Evaluating this progress in a scientifically robust approach and peer-reviewed format will be key, particularly as additional years of ASO data become available in Colorado. Research questions might, for example, include: does re-calibration of hydrologic models for improved snow process representation lead to improvements in streamflow simulations when assimilating in ASO snow data? Or, can challenges with direct insertion of ASO snow data be overcome through use of a particle filter or similar "smarter" DA techniques? Other recent scientific advancements, beyond just snow data assimilation, have also demonstrated high value for WSF improvement. CASM should stay abreast with the science and remain open to other approaches including, for example, the use of seasonal-to-subseasonal (S2S) climate forecasts (e.g., Wetterhall et al., 2016). Collaboration with researchers at NOAA's Cooperative Institute of Research to Operations in Hydrology (CIROH) may be one promising venue for staying up to date with the field and leveraging federal research funding. Whatever the path towards an improved WSF, it is critical that agreed-upon model benchmarks are established, and progress is displayed openly within the community. This includes standardization and coordination of forecast points and target periods (e.g., April-July vs. April-September) with established practices at federal forecasting agencies, but also other key aspects of the forecasts including probability distributions and forecast issue dates/latency times. These are all challenges that troubled WRF-Hydro forecasts in the 2022 forecast season, though some progress appears evident this year (hindcast demonstration and forecast delivery/verification being the key missing pieces). Furthermore, while real-time streamflow forecasting experiments like those currently supported by CASM are invaluable, hindcasting experiments similar to the Lynker/NCAR-CGD pilot study are another key tool for objectively measuring forecast improvements given the many changing elements of real-time forecasting systems that are difficult to control for.

Lynker recommends that CASM set the following goals and objectives for the 2024 forecasting season:

1. **Improved and more balanced engagement with stakeholders, including:**
    - Communicating both the opportunities and challenges of snow data assimilation (DA) in hydrologic models, including discussion of additional forecast improvement techniques
    - Addressing the challenges of translating improved snow data products into improved streamflow forecasts, particularly in an operational capacity (e.g., with a conceptual hydrologic/snow model that might not be calibrated to optimize snowpack representation)

2. **Continuation of the annual "CASM Streamflow Forecasting Roundtable" each fall**
    - Focusing on forecast review and verification, mid-term program planning, and government/academia research collaboration opportunities. Invitees should include:
        - Federal forecasters and hydrologists (e.g., NOAA's RFCs and NRCS)
        - Experimental modeling collaborators, including researchers outside of CASM
        - Key stakeholders/forecast end users

3. **Funding of additional model hindcasting experiments targeting peer-reviewed publications:**
    - Designed to objectively measure streamflow forecast improvements over existing benchmarks, particularly when/if additional years of ASO data become available in Colorado
    - But also the evaluation of other modeling approaches (see recommendation #4)

4. **Exploration of additional modeling approaches and techniques, including:**
    - Models:
        - Other process-based, intermediate-complexity models with snow-DA (beyond direct insertion), hindcasting, and real-time forecasting capabilities (e.g., NCAR-CGD's SUMMA)
        - Machine-learning models (e.g., LSTM from Upstream Tech and others)
    - Integration/assimilation of additional information using improved DA techniques, such as:
        - Other datasets: snow albedo, snow covered area, streamflow
        - Seasonal precipitation and temperature forecasts
    - Next-generation modeling capabilities from federal forecasting agencies:
        - USDA NRCS's new AI-based water supply forecasting system, M4
        - NOAA's NextGen National Water Model (NWM), a model-agnostic framework

5. **Standardization of forecasts with existing federal forecasting agency's practices, including:**
    - Forecast points and probability distributions
    - Target periods (April-July vs. April-Sept), and calibration datasets (naturalized vs. observed flows)
    - Forecast delivery mechanisms, including prioritization of machine-readable formats
    - Forecast latency time minimization

6. **Collaboration with NOAA's new Cooperative Institute for Research to Operations in Hydrology (CIROH) and other collaborators**
    - CIROH is a new $360M cooperative institute focused on improving hydrologic prediction

Together, these efforts will work towards ensuring that discussions with stakeholders around if and how to continue the CASM program are transparent and evidence-based, and that under the premise of a continued CASM program, 1) ASO-informed seasonal streamflow forecasts provide quantifiable performance benefits over currently available forecasts relative to observations; and 2) models, methods, and techniques identified for CASM experimental streamflow forecasts during 2024 and beyond are well-suited for the needs of Colorado stakeholders going forward.

7. References

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S. R. I. N. I. V. A. S. A. N. B., Srinivasan, R., Yang, H., & Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of hydrology, 524, 733-752.

Bearup, L., Gutmann, E., Wood, A., Painter, T., & Nowak, K. (2021, December). Merging Airborne Snow Data with Long-term Hydrometeorological Observations to Improve Water Supply Forecasting. In *AGU Fall Meeting Abstracts* (Vol. 2021, pp. H55F-0809).

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... & Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498-2514.

Cooperative Institute for Research to Operations in Hydrology. (2023). University of Alabama. CIROH. https://ciroh.ua.edu/

Essery, R. (2013). Snowpack modelling and data assimilation. ECMWF-WWRP/THORPEX Workshop on Polar Prediction. Available online: https://www.ecmwf.int/sites/default/files/elibrary/2013/13948-snowpackmodelling-and-data-assimilation. pdf (accessed on 13 December 2018).

Fleming, S. W., Vesselinov, V. V., & Goodbody, A. G. (2021). Augmenting geophysical interpretation of data-driven operational water supply forecast modeling for a western US river using a hybrid machine learning approach. *Journal of Hydrology*, *597*, 126327.

Harpold, A. A., Dettinger, M., & Rajagopal, S. (2017). Defining snow drought and why it matters, EOS, 98. *Eos*, *98*.

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., ... & Painter, T. H. (2018). Direct insertion of NASA Airborne Snow Observatory-derived snow depth time series into the iSnobal energy balance snow model. Water Resources Research, 54(10), 8045-8063.

Lahmers, T. M., Kumar, S. V., Rosen, D., Dugger, A., Gochis, D. J., Santanello, J. A., ... & Dunlap, R. (2022). Assimilation of NASA's Airborne Snow Observatory Snow Measurements for Improved Hydrological Modeling: A Case Study Enabled by the Coupled LIS/WRF-Hydro System. *Water Resources Research*, *58*(3), e2021WR029867.

Lehner, F., Wood, A. W., Llewellyn, D., Blatchford, D. B., Goodbody, A. G., & Pappenberger, F. (2017). Mitigating the impacts of climate nonstationarity on seasonal streamflow predictability in the US Southwest. *Geophysical Research Letters*, *44*(24), 12-208.

Lins, H. F., & Slack, J. R. (1999). Streamflow trends in the United States. Geophysical research letters, 26(2), 227-230.

Livneh, B., & Badger, A. M. (2020). Drought less predictable under declining future snowpack. *Nature Climate Change*, *10*(5), 452-458.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R., & Jonas, T. (2017). Improving physically based snow simulations by assimilating snow depths using the particle filter. Water Resources Research, 53(2), 1125-1143.

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 operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *21*(7), 3915-3935.

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., ... & Brekke, L. D. (2016). mizuRoute version 1: a river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, *9*(6), 2223-2238.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., ... & Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481-2500.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., & Nearing, G. (2017). Benchmarking of a physically based hydrologic model. *Journal of Hydrometeorology*, *18*(8), 2215-2225.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., ... & Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. Hydrology and Earth System Sciences, 26(15), 4013-4032.

NOAA (2022). *CBRFC Water Year in Review An Overview of Operational Changes, Improvements, and Investigations over the course of Water Year 2021*. CBRFC. https://www.cbrfc.noaa.gov/present/yir/yir_2021.pdf

Pagano, T., & Garen, D. (2005). A recent increase in western US streamflow variability and persistence. *Journal of Hydrometeorology*, *6*(2), 173-179.

Pagano, T., Garen, D., & Sorooshian, S. (2004). Evaluation of official western US seasonal water supply outlooks, 1922–2002. *Journal of Hydrometeorology*, *5*(5), 896-909.

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward earlier streamflow timing across western North America. Journal of climate, 18(8), 1136-1155.

Sturtevant, J., Wood, A., Llewellyn, D., Barrett, L., & Lehner, F. (2021, December). Assessing seasonal streamflow forecast strategies for Upper Rio Grande basin water management using SUMMA watershed models. In *AGU Fall Meeting Abstracts* (Vol. 2021, pp. H45V-1453).

Sturtevant, J., Wood, A. W., Llewellyn, D., Barrett, L. J., Lehner, F., Mizukami, N., & Liu, H. (2020, December). Improving the robustness of Upper Rio Grande water supply forecasting in the face of climate trends and variability. In *AGU Fall Meeting Abstracts* (Vol. 2020, pp. H041-03).

Thirel, G., Salamon, P., Burek, P., & Kalas, M. (2013). Assimilation of MODIS snow cover area data in a distributed hydrological model using the particle filter. Remote Sensing, 5(11), 5825-5850.

Wetterhall, F., Pechlivanidis, I. G., Ramos, M. H., Wood, A. W., Wang, Q. J., Zehe, E., & Ehret, U. (2016). Sub-seasonal to seasonal hydrological forecasting. *Hydrology and Earth System Sciences*.

Wood, A., Liu, H., Lanini, J., McGuire, M., Broman, D., Erger, P., & Bearup, L. (2021a). Using multi-scale SUMMA watershed modeling for seasonal streamflow prediction to support water operations at Buffalo Bill Reservoir, Wyoming. In *AGU Fall Meeting Abstracts* (Vol. 2021, pp. H42F-06).

Wood, A. W., Mizukami, N., Clark, M. P., Newman, A. J., Liu, H., Sturtevant, J., ... & Arnold, J. R. (2020). A new SUMMA and MizuRoute hydrologic modeling resource for US water applications. In *AGU Fall Meeting Abstracts* (Vol. 2020, pp. H041-02).

Wood, A. W., Sturtevant, J. T., Barrett, L., & Llewellyn, D. (2021b). *Improving the reliability of southwestern US water supply forecasting*. Research and Development Office. https://www.usbr.gov/research/projects/download_product.cfm?id=3029

Wood, A. W., Sturtevant, J., Frans, C. D., Mizukami, N., Nijssen, B., & Harrell, J. (2022). A new process-oriented ensemble hydrological prediction system for flood prediction and water management in the US Pacific Northwest. In *AGU Fall Meeting Abstracts* (Vol. 2022, pp. H15D-08).

8. Appendices

8.1. Hydroclimate Data

8.1.1. Taylor River at Taylor Park

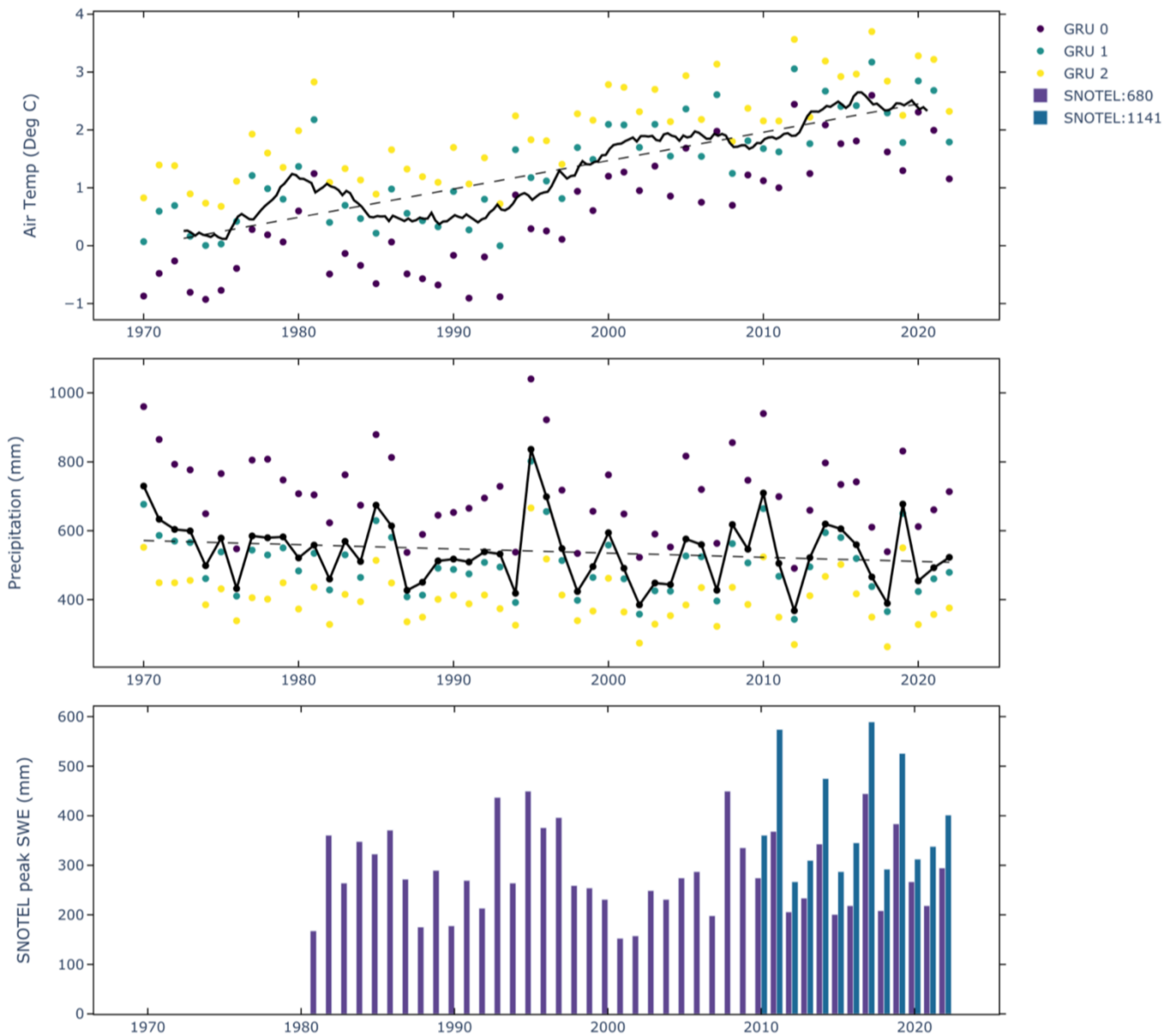


Figure 11: Taylor River at Taylor Park forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.1.2. Willow Creek at Willow Creek Reservoir

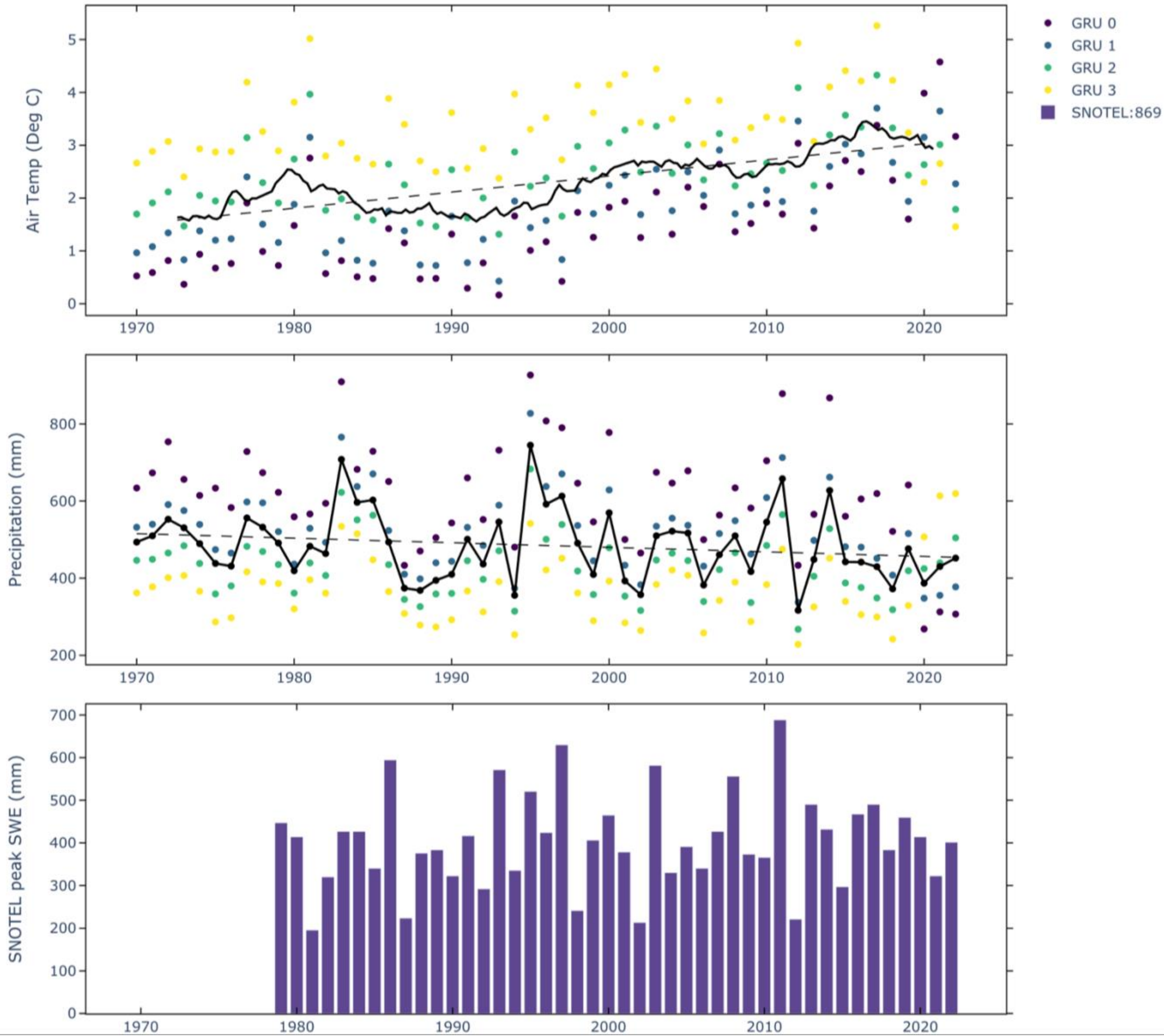


Figure 12: Willow Creek forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.1.3. Lake Granby

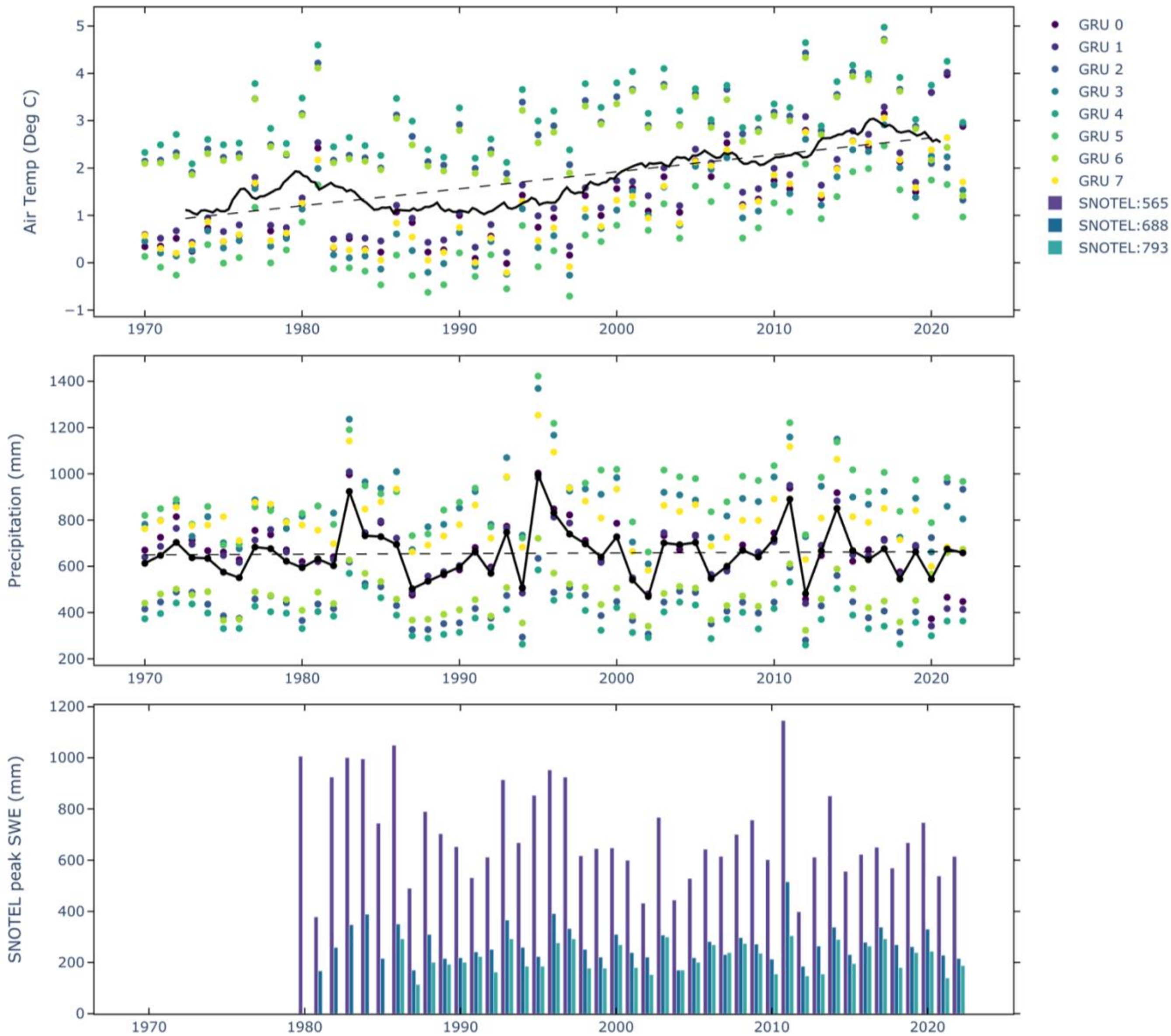


Figure 13: Lake Granby forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.1.4. Dillon Reservoir

Figure 14: Dillon Reservoir forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.1.5. Dolores River at Dolores

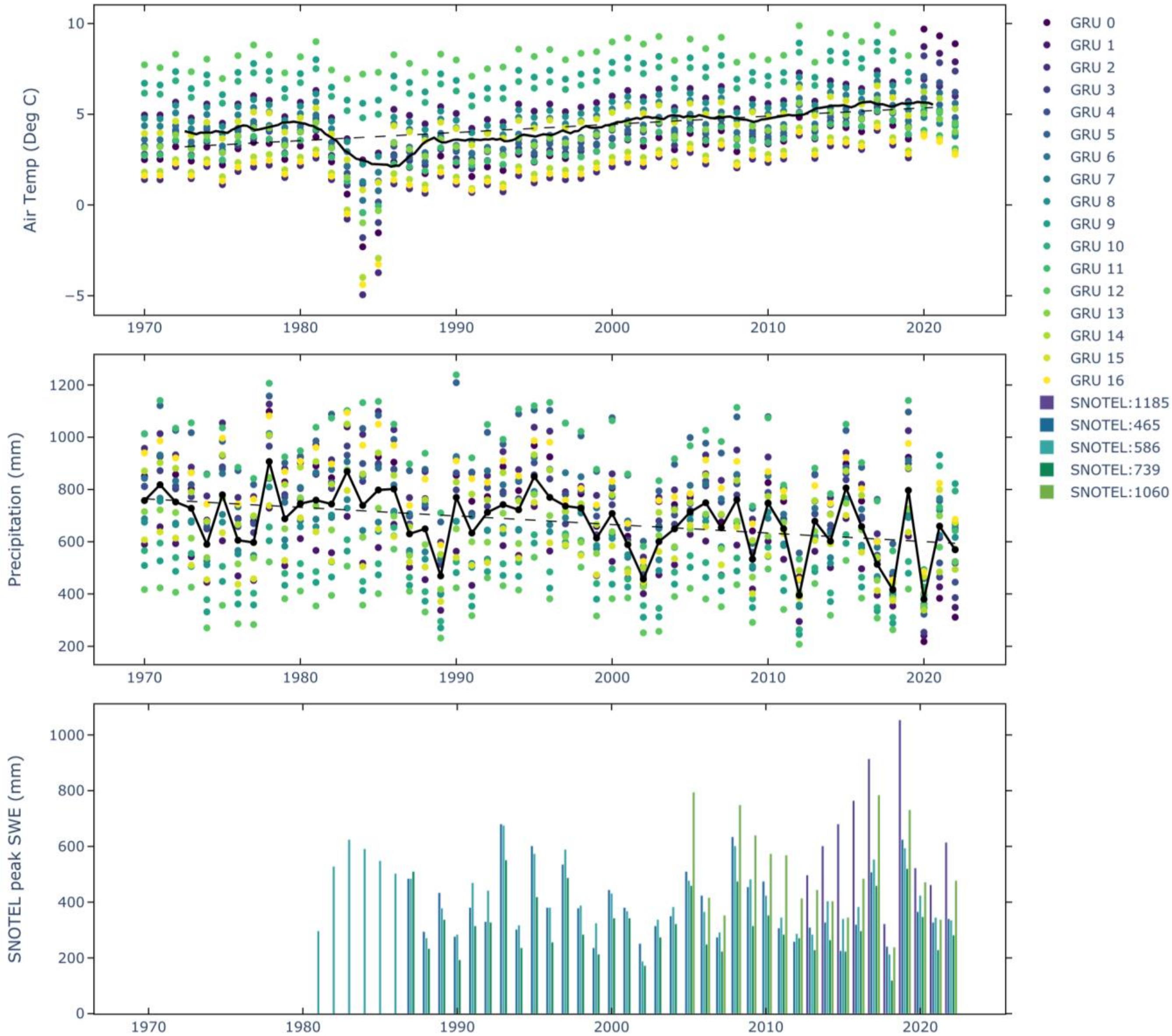


Figure 15: Dolores River forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel). The dip in mid-1980s air temperature is likely an artifact of a new station coming online, which affects the underlying dataset.

8.1.6. Fraser River at Winter Park

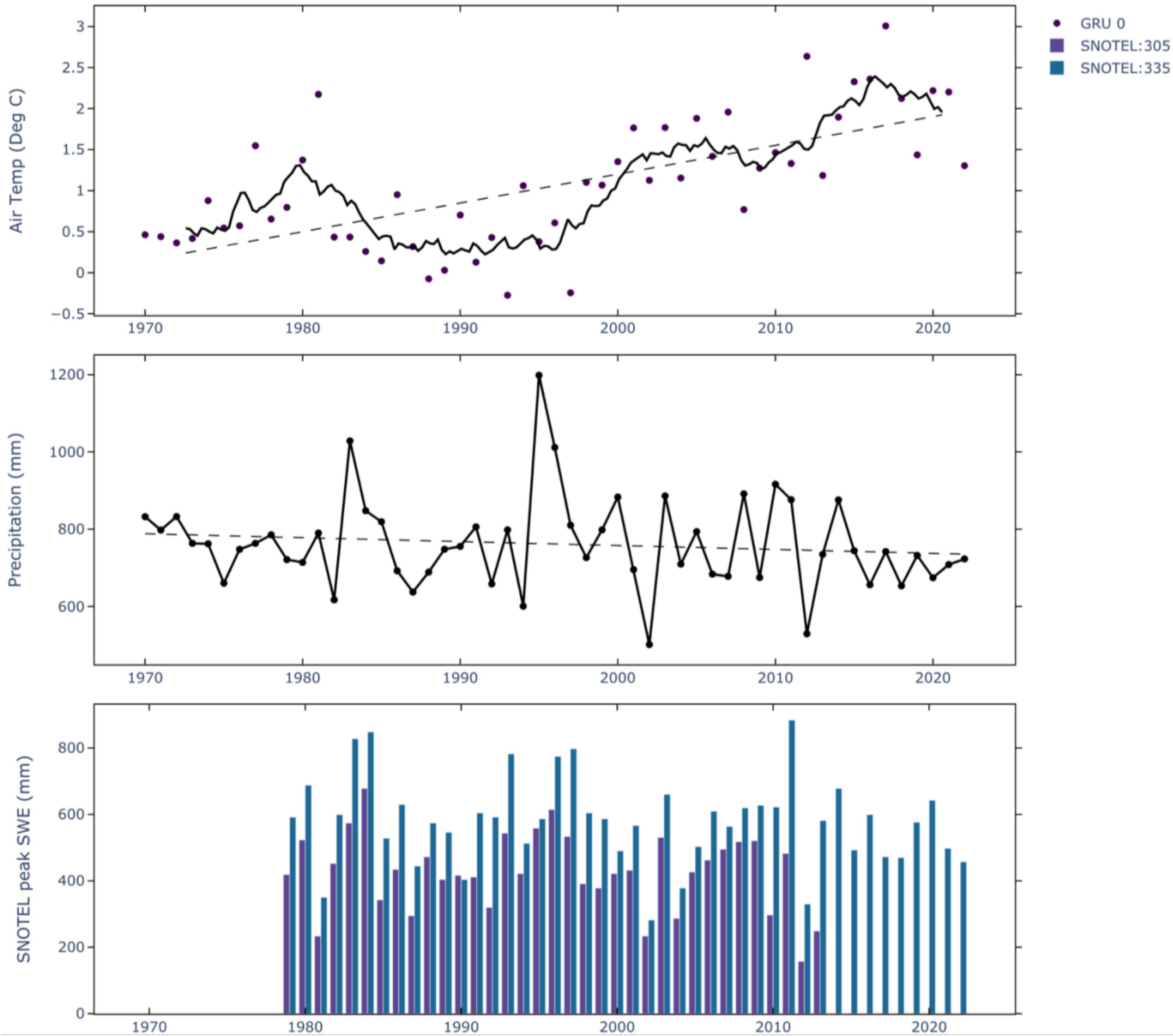


Figure 16: Fraser River at Winter Park forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.1.7. Conejos River near Mogote

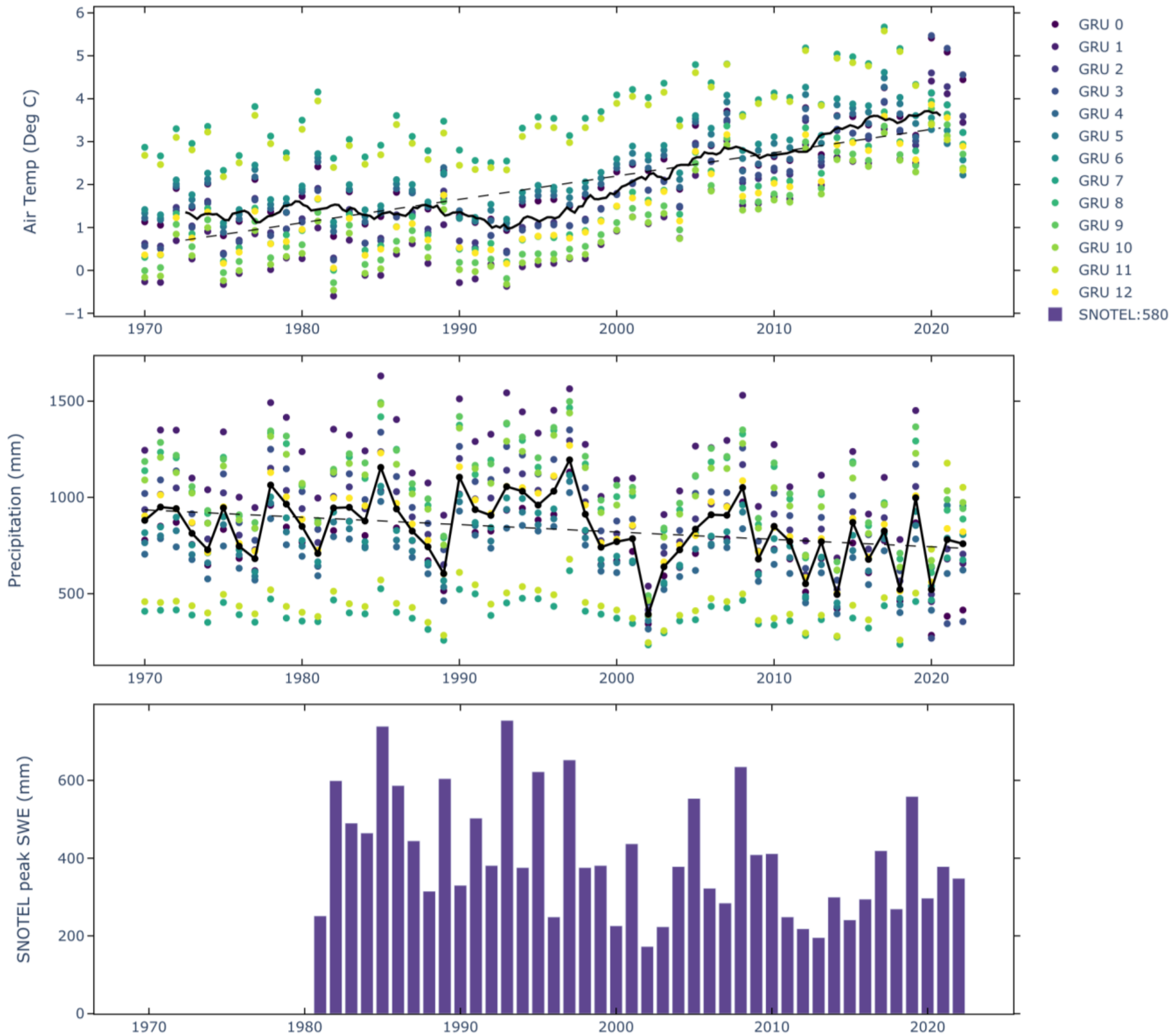


Figure 17: Conejos River near Mogote forcing data and hydroclimate trends from the GMET dataset and the nearest SNOTEL station. Variables include air temperature (GMET; top panel), precipitation (GMET; middle panel), and peak snowpack (SNOTEL; bottom panel).

8.2. Historical Water Supply Forecasts

8.2.1. Taylor River at Taylor Park

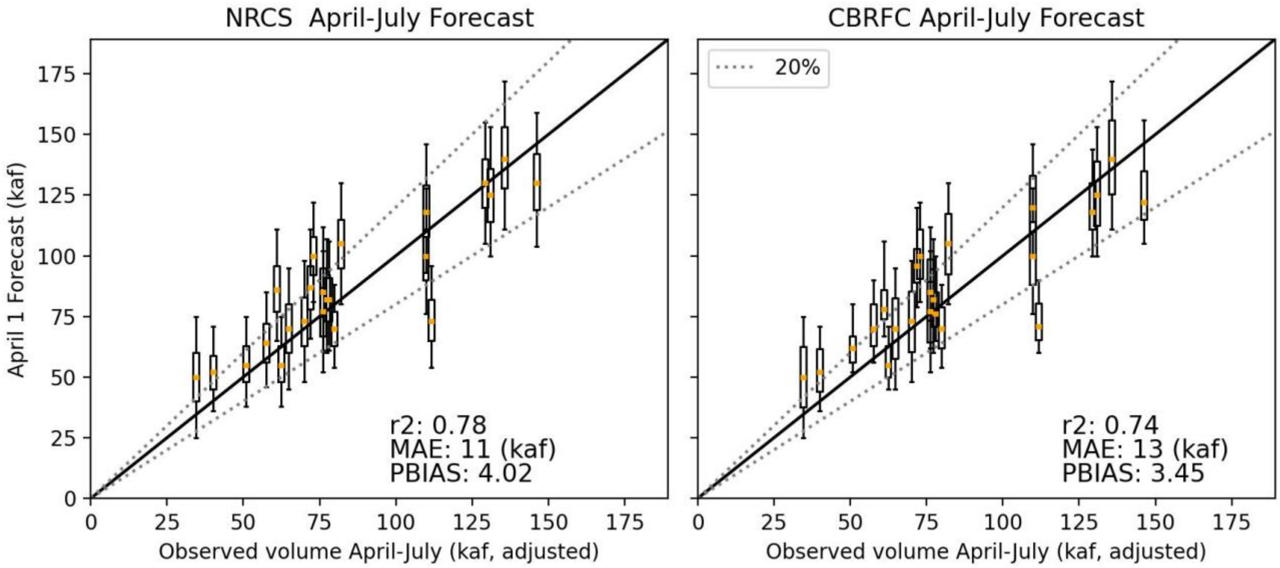


Figure 18: Taylor River at Taylor Park April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume.

8.2.2. Willow Creek at Willow Creek Reservoir

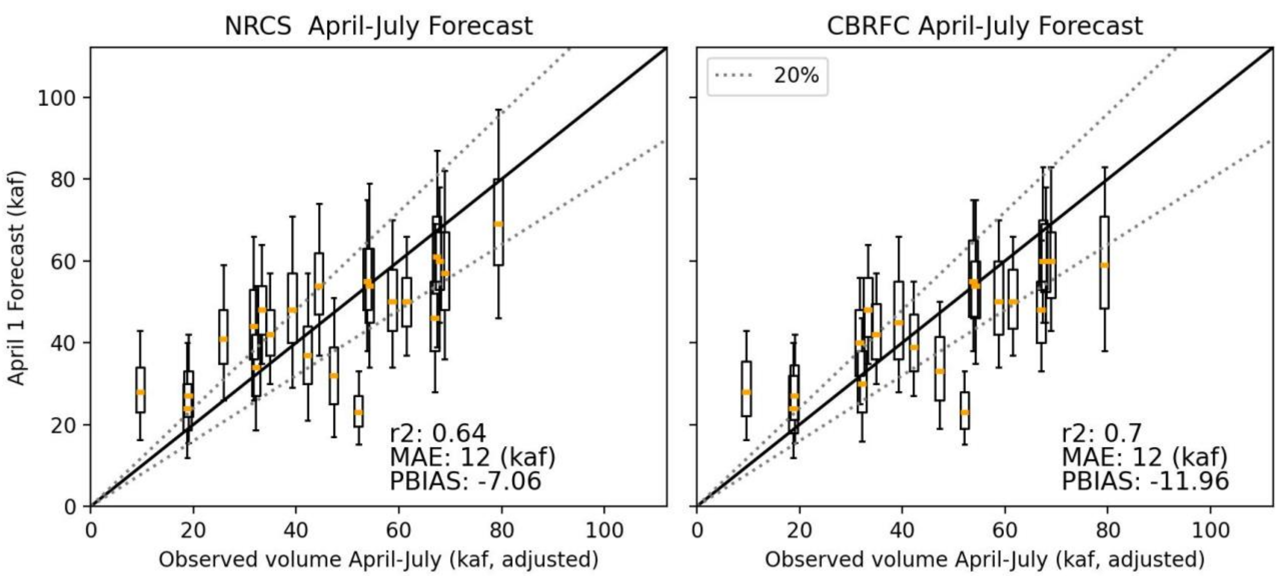


Figure 19: Willow Creek Reservoir April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume.

8.2.3. Lake Granby

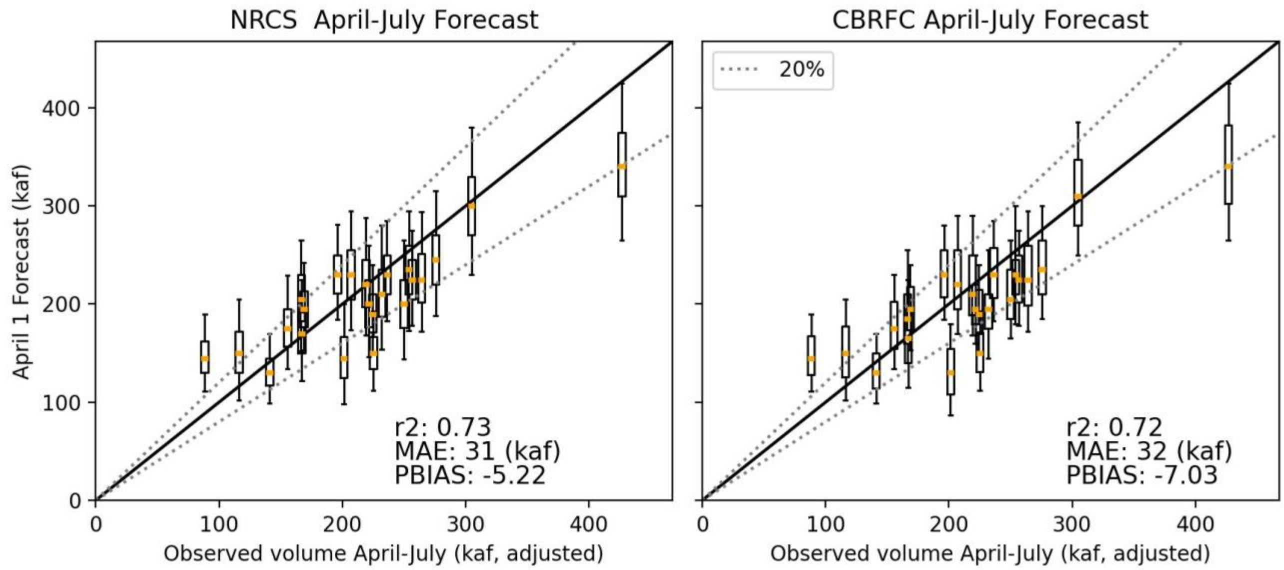


Figure 20: Lake Granby April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume.

8.2.4. Dillon Reservoir

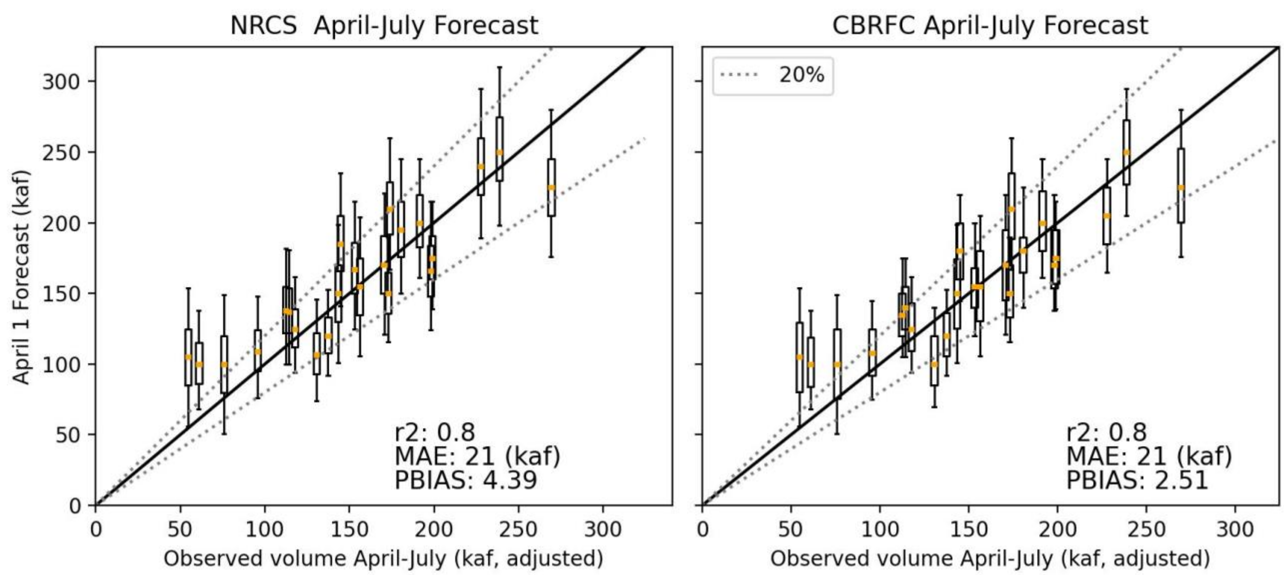


Figure 21: Dillon Reservoir April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume.

8.2.5. Dolores River at Dolores

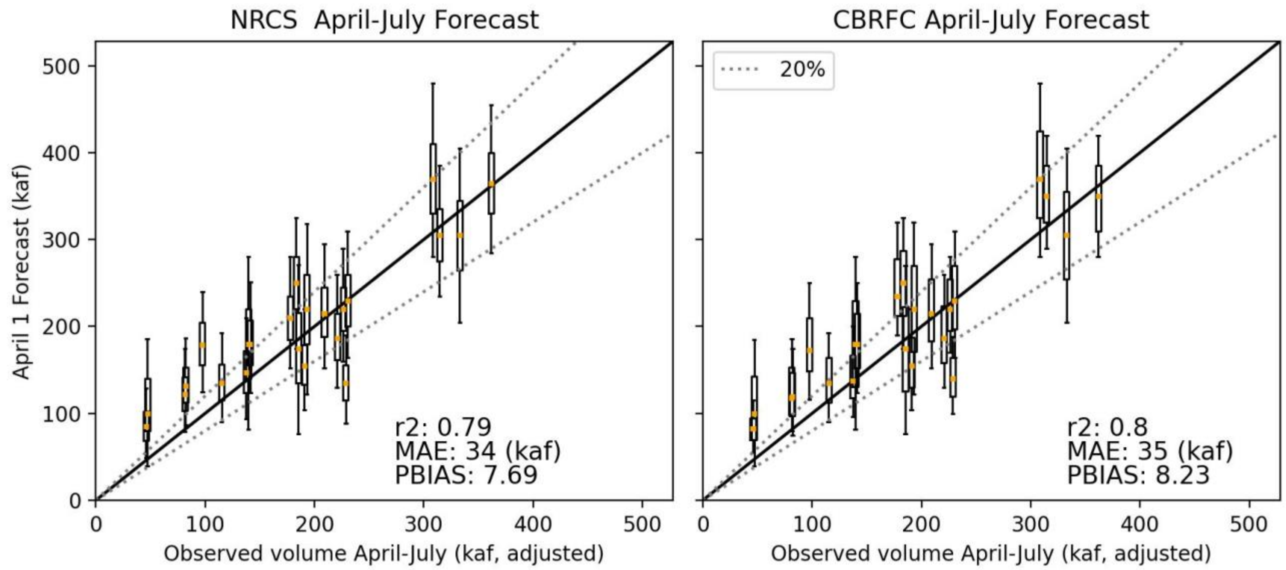


Figure 22: Dolores River at Dolores April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from NRCS (left) and CBRFC (right). Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume.

8.2.6. Fraser River at Winter Park

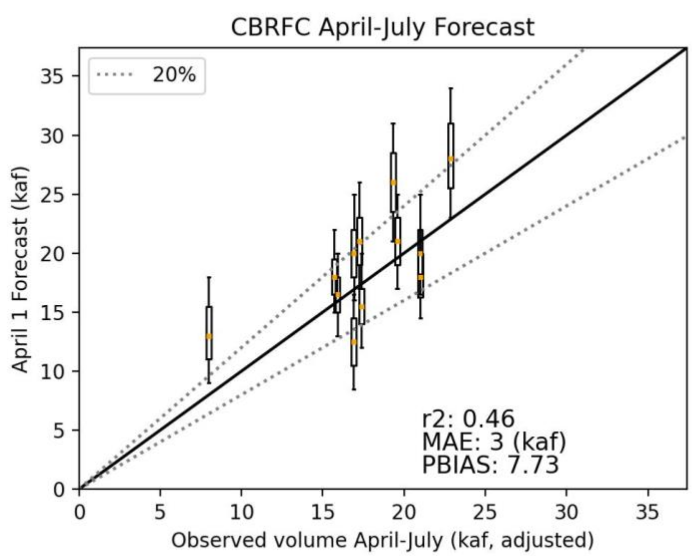


Figure 23: Fraser River at Winter Park April 1st forecasts for April-July (AMJJ) volumes vs. observed AMJJ streamflow from CBRFC. Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. The period of record is limited by the lack of observed data for the Jim Creek Diversion from CO DWR, which is used to calculate the adjusted April-July volume of the Fraser River at Winter Park.

8.2.7. Conejos River near Mogote

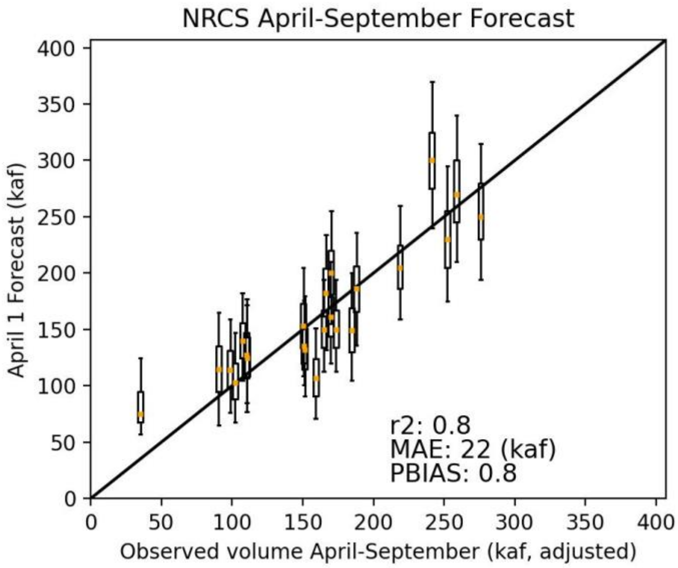


Figure 24: Conejos River near Mogote April 1st forecasts for April-September (AMJJAS) volumes vs. observed AMJJAS streamflow from NRCS. Whiskers denote the 10% and 90% exceedance levels, box edges show 30% and 70% levels, and the orange dot notes the 50% level (median). Dashed lines represent the 20% bounds on the observed volume. WGRFC does forecast for this point, however the period of record is limited and thus is excluded from the study.

8.3. Experimental Forecast Documentation

8.3.1. Water Supply Forecast Figures

8.3.1.1. Taylor River at Taylor Park

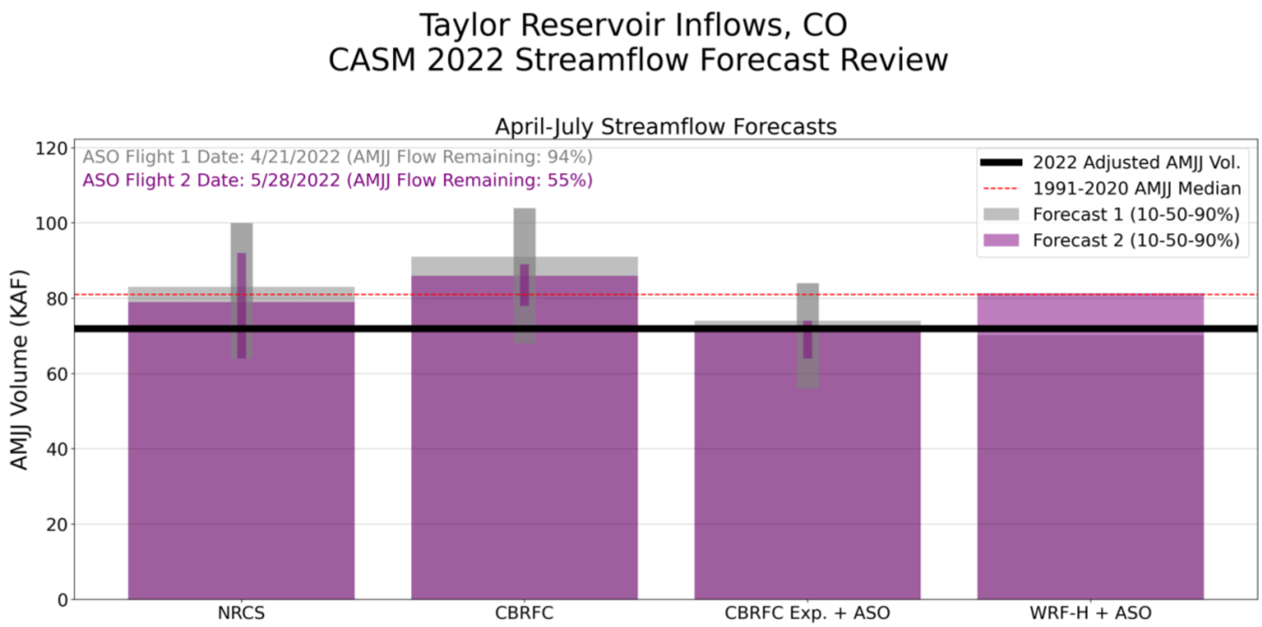


Figure 25: Taylor Reservoir official and experimental water supply forecasts. SUMMA hindcasts are not available for Taylor Reservoir, since the model was run for the Taylor River at Taylor Park forecast point just upstream of the reservoir. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing.

8.3.1.2. Willow Creek at Willow Creek Reservoir

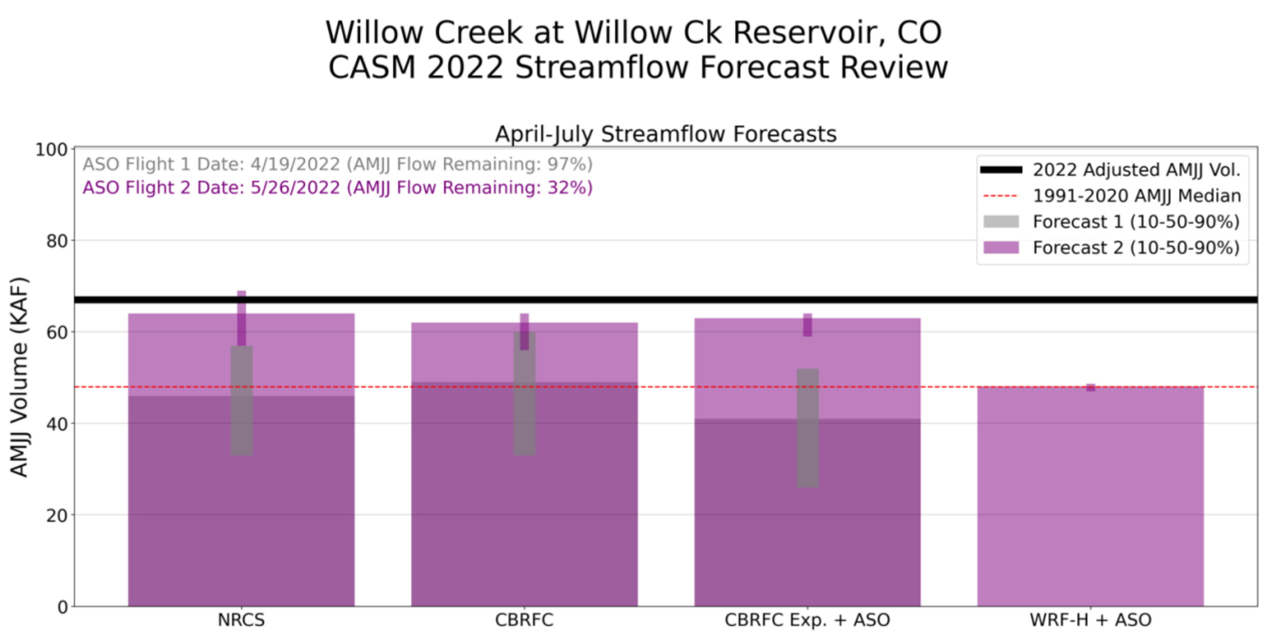


Figure 26: Willow Creek at Willow Creek Reservoir official and experimental water supply forecasts. Forecast 1 from WRF-Hydro was not issued due to technical issues of the forecasting system. Additionally, WRF-Hydro forecasts were issued as natural flows into a lake with a fill-and-spill type operation, which did not explicitly represent the operations of Willow Creek Reservoir.

8.3.1.3. Lake Granby

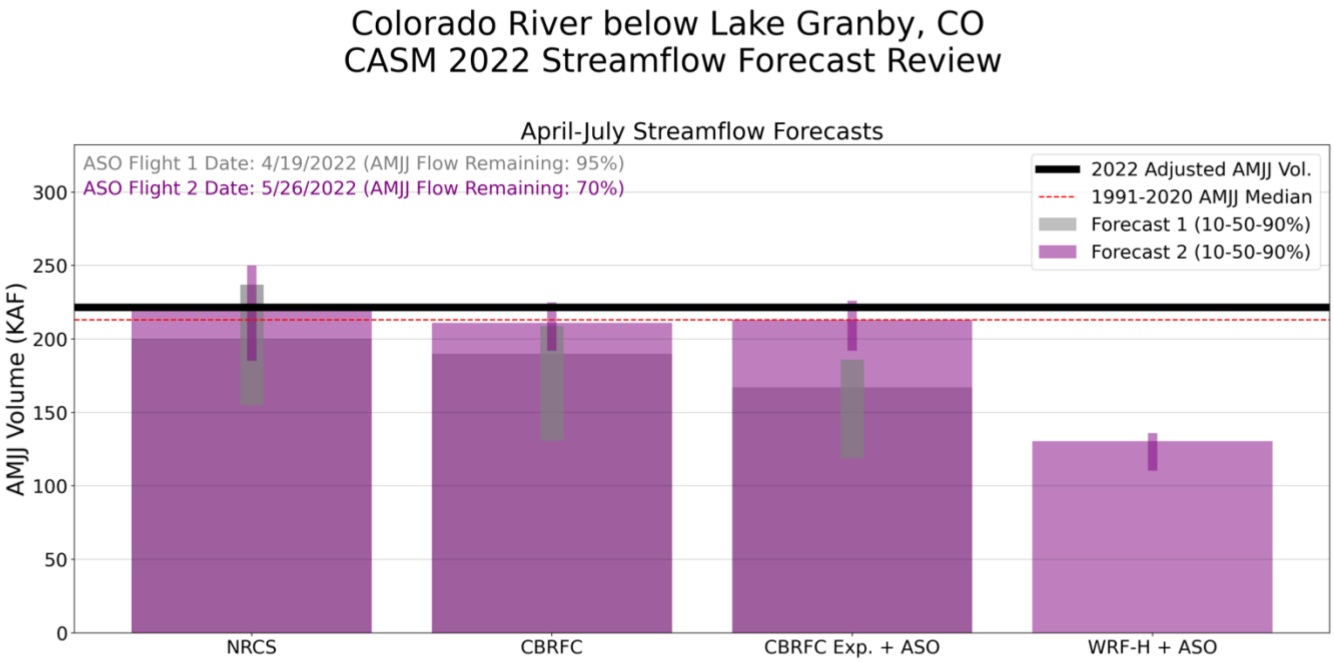


Figure 27: Lake Granby official and experimental water supply forecasts. WRF-Hydro forecasts were issued as natural flows into a lake with a fill-and-spill type operation, which poorly simulated the NRCS-reported 2022 AMJJ adjusted volumes. Additionally, Forecast 1 from WRF-Hydro was not issued due to technical issues of the forecasting system.

8.3.1.4. Dillon Reservoir

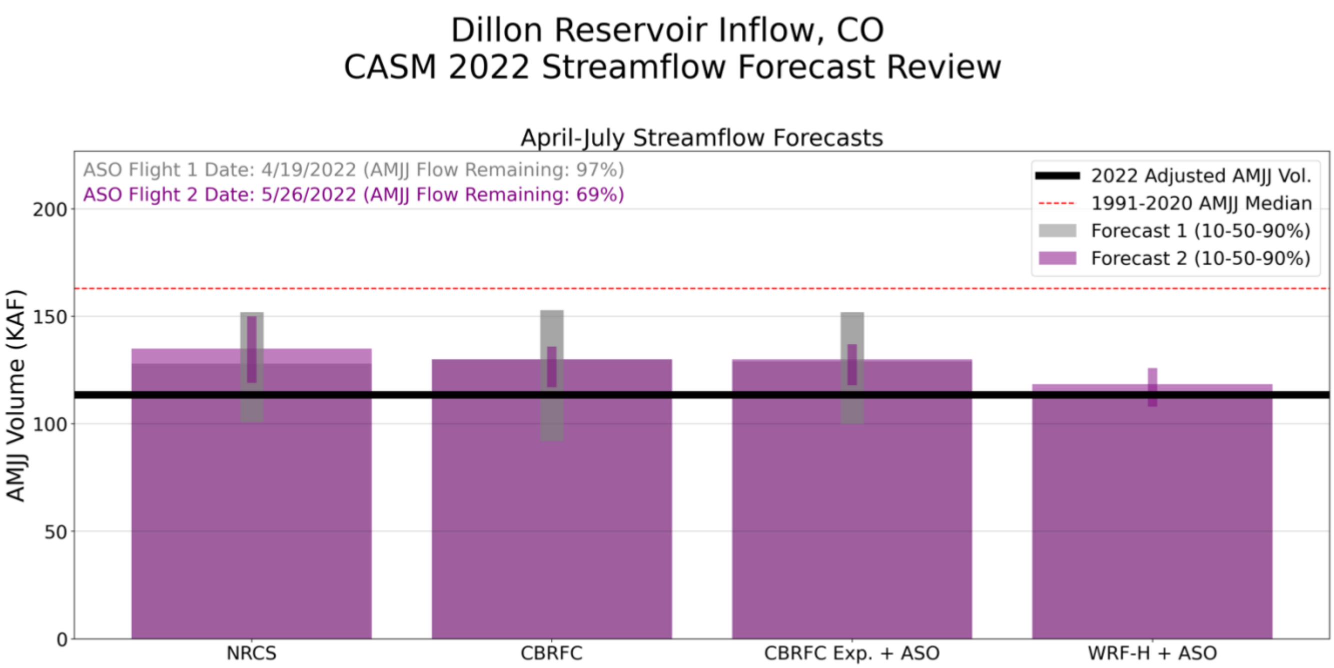


Figure 28: Dillon Reservoir official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing.

8.3.1.5. Dolores River at Dolores

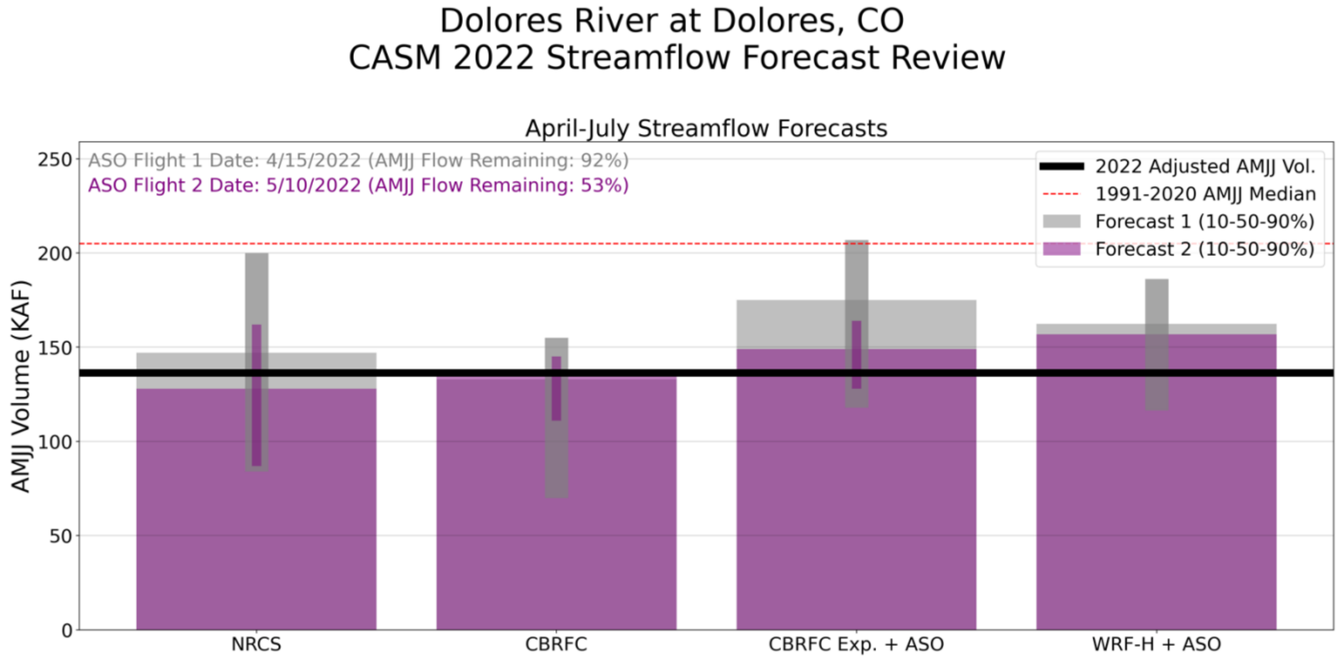


Figure 29: Dolores River at Dolores official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing.

8.3.1.6. Fraser River at Winter Park

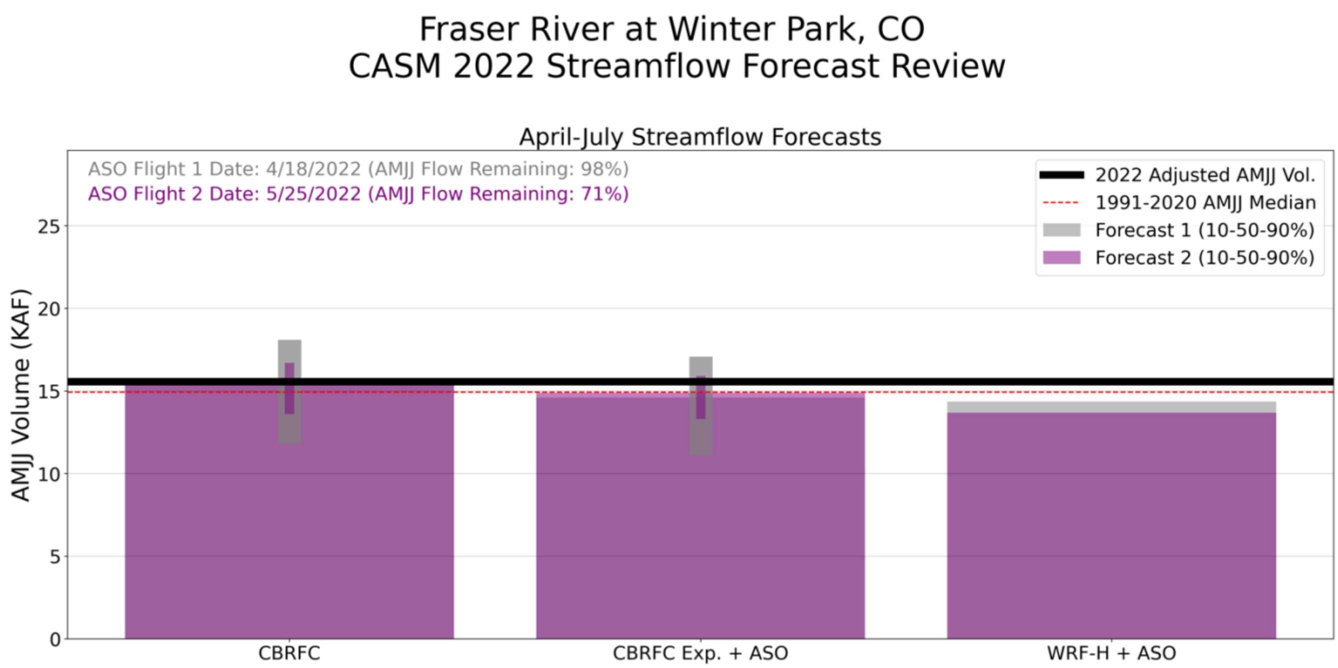


Figure 30: Fraser River at Winter Park official and experimental water supply forecasts. Forecast 1 from WRF-Hydro is for an April-September period; August-September average flows (NRCS) were removed in post-processing.

8.3.1.7. Conejos River near Mogote

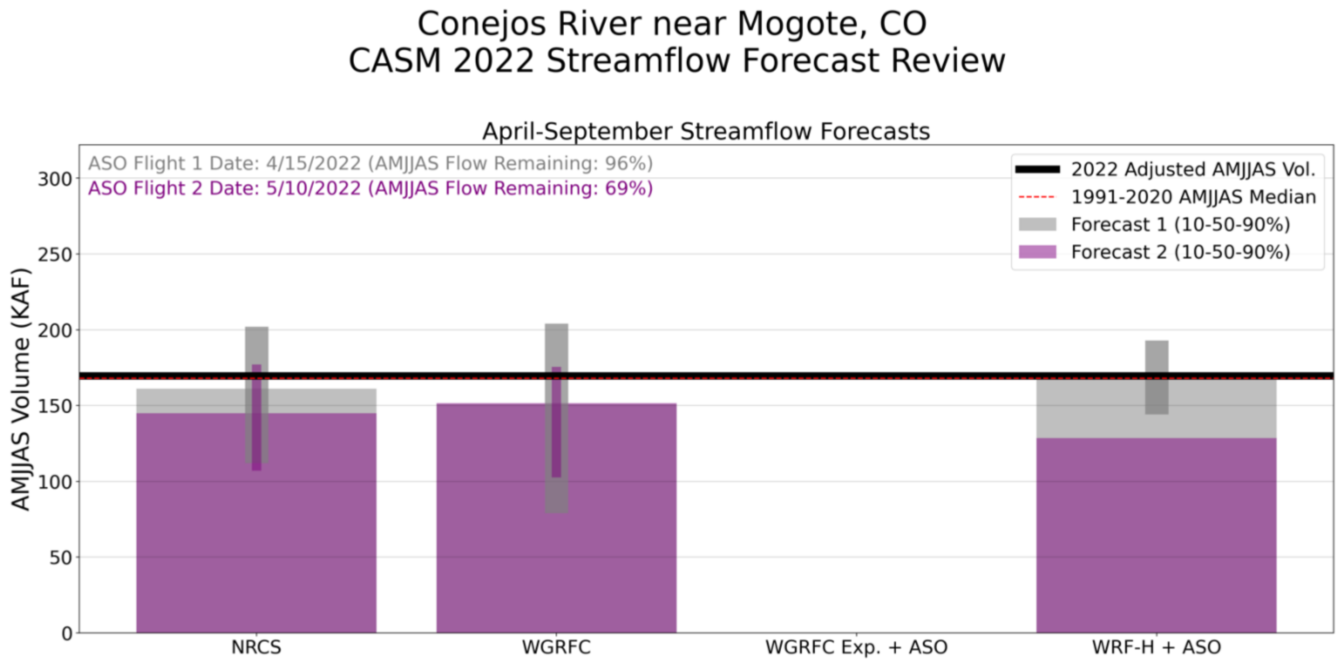


Figure 31: Conejos River at Mogote official and experimental water supply forecasts; note that the Conejos River forecast target period is April-September, not April-July.

8.3.2. WRF-Hydro Experimental Forecasts- Tabulated

Table 6: WRF-Hydro experimental water supply forecasts, flight one. All volumes are in KAF.

Basin ID	Forecast Point Name	River Basin	AMJJ Observed	10%	30%	50%	70%	90%	Flight Date
1	East River at Almont	East	151.19	-	-	176.28*	-	-	April 21st, 2022
2	Taylor Park Reservoir Inflows	Taylor	71.89	-	-	70.22*	-	-	April 21st, 2022
3	Willow Creek at Willow Creek Reservoir	Colorado	66.98	-	-	-	-	-	April 19th, 2022
4	Lake Granby Inflows	Colorado	221.23	-	-	-	-	-	April 19th, 2022
5	Dillon Reservoir Inflows	Blue	112.49	-	-	112.04*	-	-	April 19th, 2022
6	Dolores River at Dolores	Dolores	137.43	208*	-	162.3*	-	138.4*	April 15th, 2022
7	Fraser River at Winter Park	Fraser	15.56	-	-	14.37*	-	-	April 18th, 2022
8	Conejos River near Mogote	Conejos	169.71**	192.1**	-	168.2**	-	143.4**	April 15th, 2022

**An April-September volume forecast was issued by WRF-Hydro when an April-July forecast is operationally traditional. To estimate an AMJJ volume consistent with the forecast target period used by official forecasts, average August-September flows (from NRCS) were removed from these WRF-Hydro forecasts. Missing data (e.g., Lake Granby) are due to technical issues of the forecasting system, or incomplete information on the full forecast probability distributions.*

***The Conejos at Mogote is traditionally forecasted for April-September*

Table 7: WRF-Hydro experimental water supply forecasts, flight two. All volumes are in KAF.

Basin ID	Forecast Point Name	River Basin	AMJJ Observed	10%	30%	50%	70%	90%	Flight Date
1	East River at Almont	East	151.19	175	-	163	-	158	May 28th, 2022
2	Taylor Park Reservoir Inflows	Taylor	71.89	-	-	81.3	-	-	May 28th, 2022
3	Willow Creek at Willow Creek Reservoir	Colorado	66.98	49.2	-	48.1	-	47.5	May 26th, 2022
4	Lake Granby Inflows	Colorado	221.23	150.7	-	130.5	-	125.1	May 26th, 2022
5	Dillon Reservoir Inflows	Blue	112.49	129	-	118.5	-	111	May 26th, 2022
6	Dolores River at Dolores	Dolores	137.43	-	-	156.9	-	-	May 10th, 2022
7	Fraser River at Winter Park	Fraser	15.56	-	-	13.7	-	-	May 25th, 2022
8	Conejos River near Mogote	Conejos	169.71**	-	-	128.4**	-	-	May 10th, 2022

**An April-September volume forecast was issued by WRF-Hydro when an April-July forecast is operationally traditional. To estimate an AMJJ volume consistent with the forecast target period used by official forecasts, average August-September flows (from NRCS) were removed from these WRF-Hydro forecasts. Missing data (e.g., Lake Granby) are due to technical issues of the forecasting system, or incomplete information on the full forecast probability distributions.*

***The Conejos at Mogote is traditionally forecasted for April-September*

8.3.3. CBRFC Experimental Forecasts- Tabulated

Table 8: CBRFC experimental (ASO SWE-data assimilation) water supply forecasts, flight one. All volumes are in KAF.

Basin ID	Forecast Point Name	River Basin	AMJJ Observed	10%	30%	50%	70%	90%	Flight Date
1	East River at Almont	East	151.19	131	109	99	91	83	April 21st, 2022
2	Taylor Park Reservoir Inflows	Taylor	71.89	92	78	74	69	64	April 21st, 2022
3	Willow Creek at Willow Creek Reservoir	Colorado	66.98	56	44	41	35	30	April 19th, 2022
4	Lake Granby Inflows	Colorado	221.23	215	187	167	159	148	April 19th, 2022
5	Dillon Reservoir Inflows	Blue	112.49	158	140	129	117	106	April 19th, 2022
6	Dolores River at Dolores	Dolores	137.43	232	204	175	162	143	April 15th, 2022
7	Fraser River at Winter Park	Fraser	15.56	18.1	15.8	14.6	13.7	12.1	April 18th, 2022
8	Conejos River near Mogote	Conejos	169.71*	-	-	-	-	-	April 15th, 2022

**The Conejos at Mogote is traditionally forecasted for April-September*

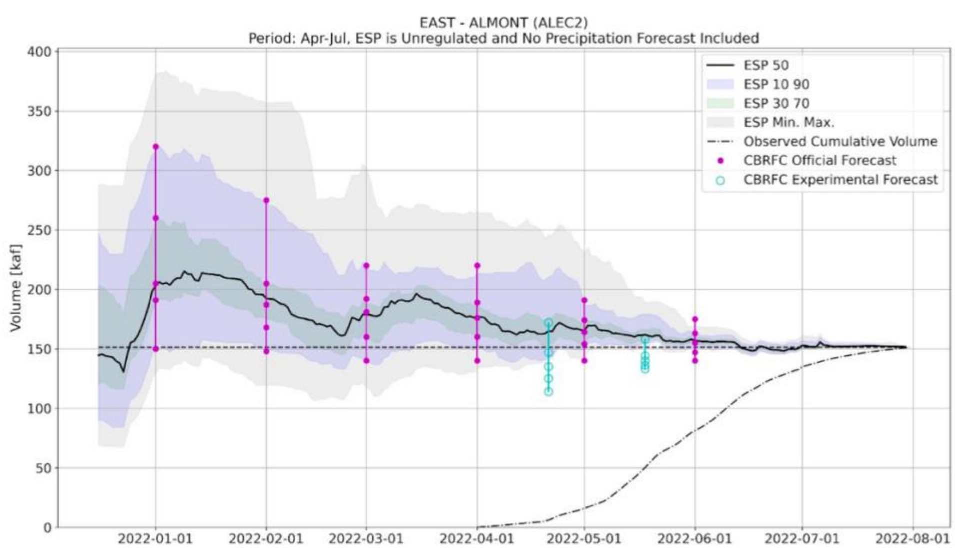


Figure 32: An example of CBRFC water supply forecasts for 2022, where official forecast guidance is in purple compared with ASO-informed forecasts in light blue (April 21st, 2022, and May 18th, 2022)

Table 9: CBRFC experimental (ASO SWE-data assimilation) water supply forecasts, flight two. All volumes are in KAF.

Basin ID	Forecast Point Name	River Basin	AMJJ Observed	10%	30%	50%	70%	90%	Flight Date
1	East River at Almont	East	151.19	158	144	140	136	133	May 18th, 2022
2	Taylor Park Reservoir Inflows	Taylor	71.89	78	72	71	69	68	May 25th, 2022
3	Willow Creek at Willow Creek Reservoir	Colorado	66.98	67	65	63	62	62	May 26th, 2022
4	Lake Granby Inflows	Colorado	221.23	234	219	213	203	200	May 26th, 2022
5	Dillon Reservoir Inflows	Blue	112.49	142	134	130	126	123	May 26th, 2022
6	Dolores River at Dolores	Dolores	137.43	170	156	149	141	134	May 10th, 2022
7	Fraser River at Winter Park	Fraser	15.56	16.5	15.7	14.9	14.4	13.9	May 25th, 2022
8	Conejos River near Mogote	Conejos	169.71*	-	-	-	-	-	May 10th, 2022

**The Conejos at Mogote is traditionally forecasted for April-September*

8.4. Forecast Error Attribution: Spring Precipitation

8.4.1. Taylor River at Taylor Park

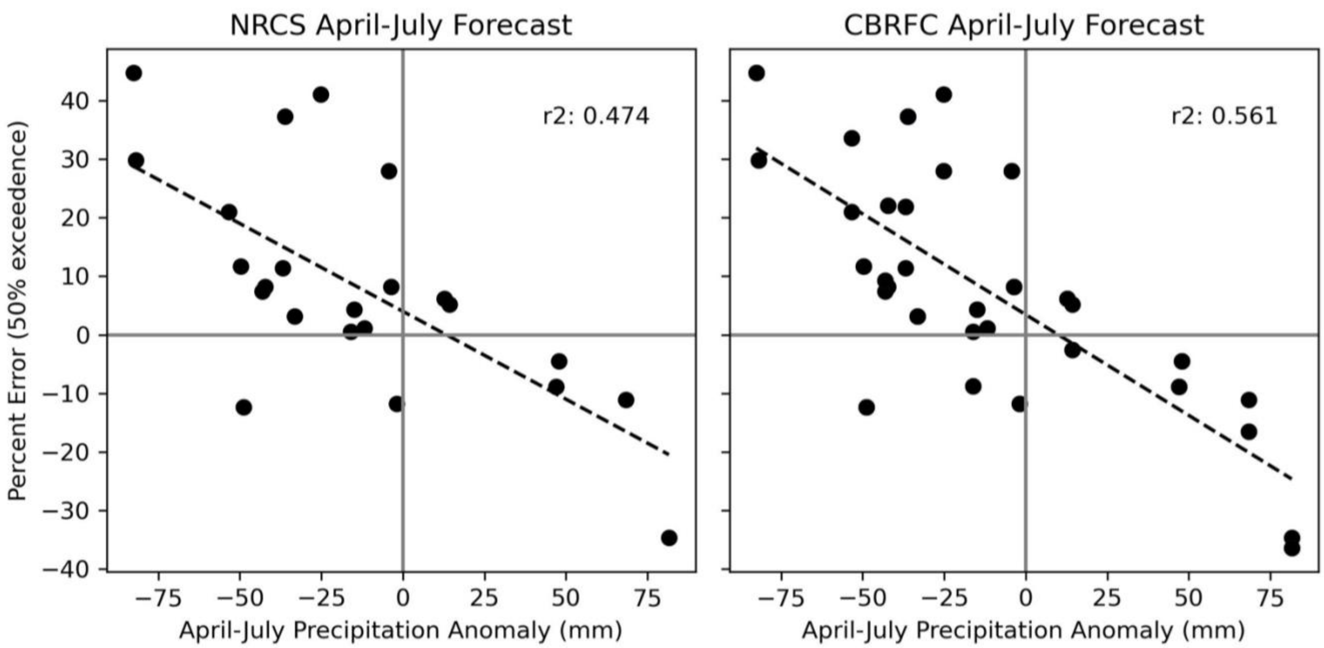


Figure 33: Taylor River at Taylor Park April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset.

8.4.2. Willow Creek at Willow Creek Reservoir

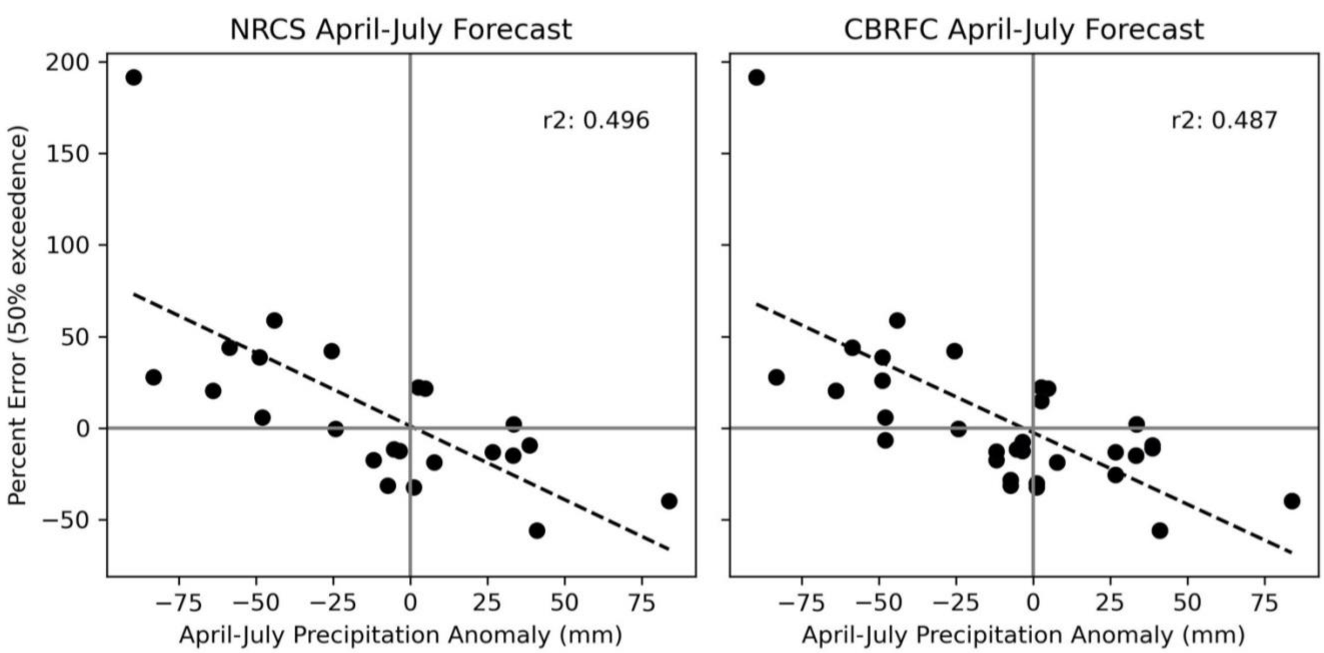


Figure 34: Willow Creek Reservoir April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset.

8.4.3. Lake Granby

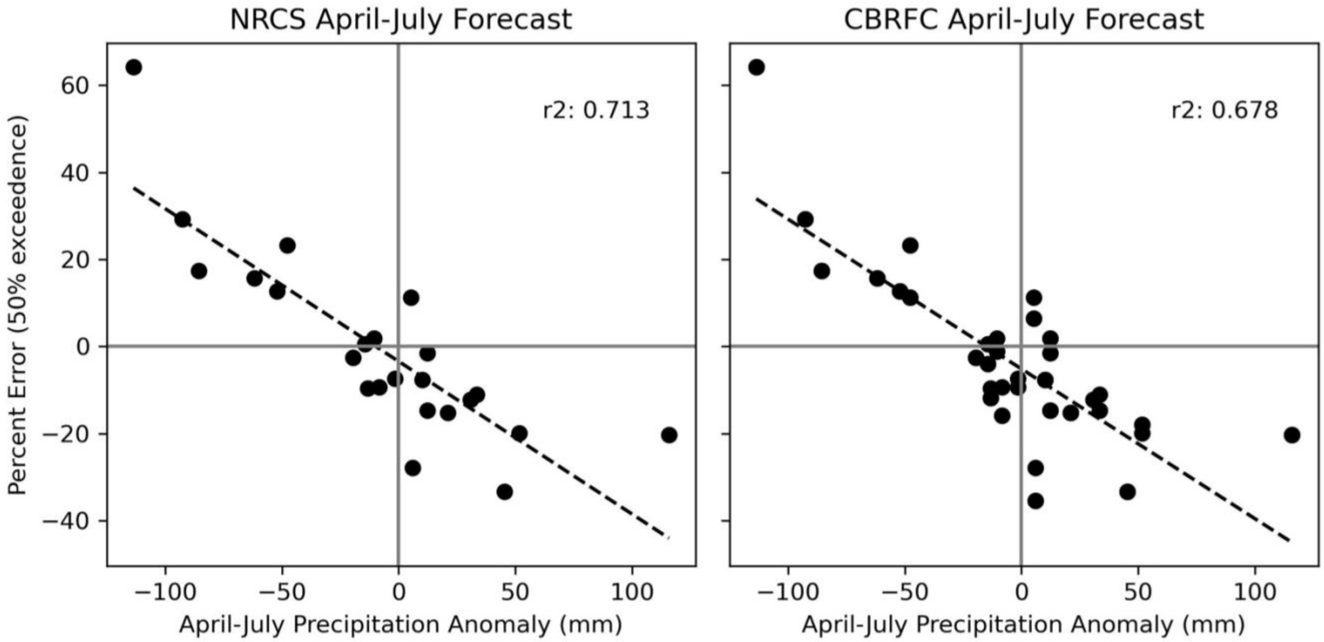


Figure 35: Lake Granby April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset.

8.4.4. Dillon Reservoir

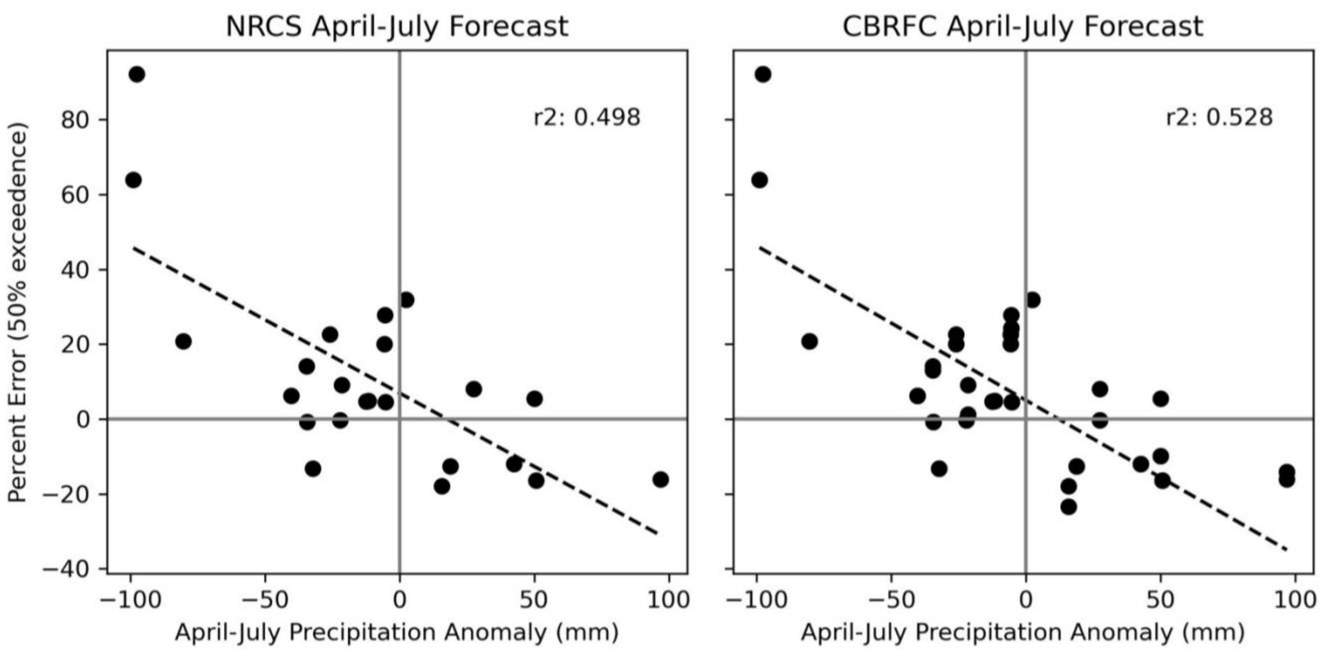


Figure 36: Dillon Reservoir April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset.

8.4.5. Dolores River at Dolores

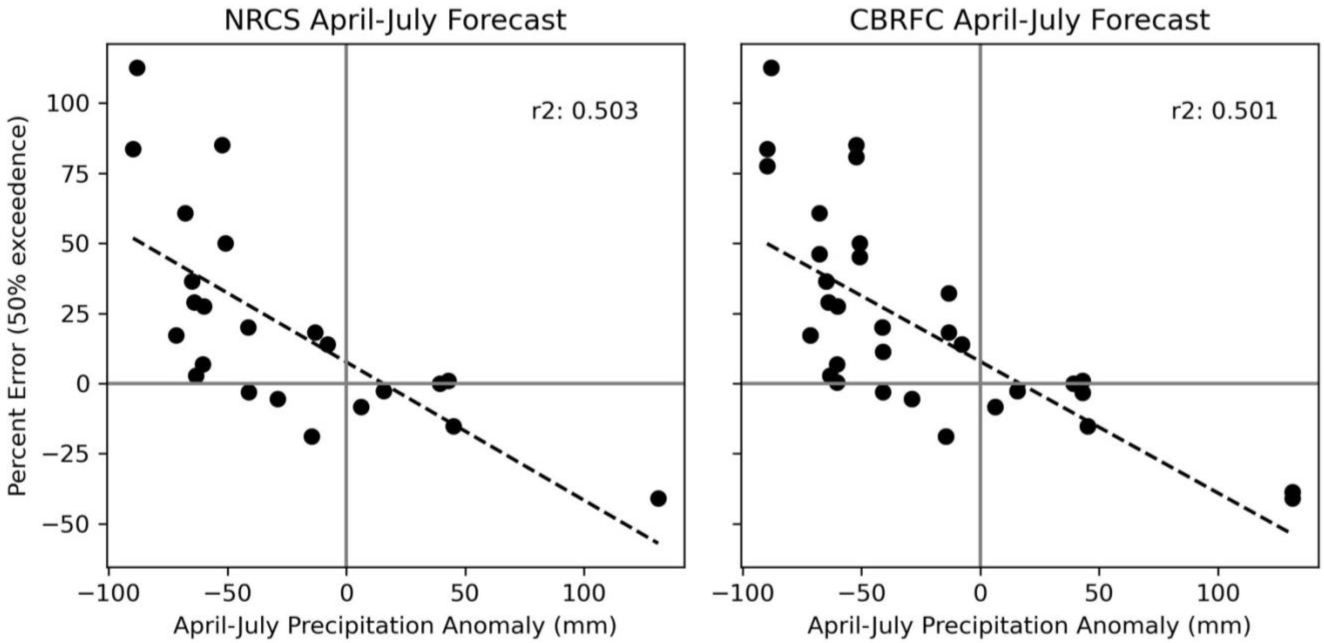


Figure 37: Dolores River at Dolores April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for NRCS (left) and CBRFC (right), as measured by the GMET forcing dataset.

8.4.6. Fraser at Winter Park

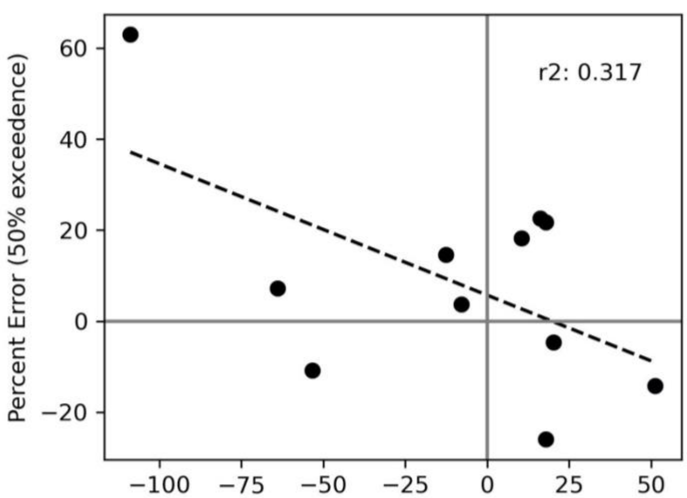


Figure 38: Fraser River at Winter Park April 1st forecast percent error (for the 50% exceedance forecast of April-July volumes) vs. April-July precipitation anomalies for CBRFC (right), as measured by the GMET forcing dataset.

8.4.7. Conejos River near Mogote

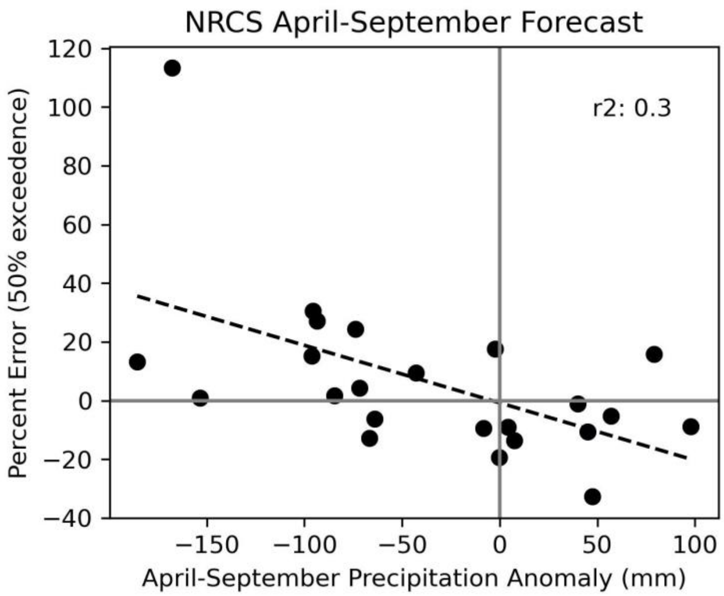


Figure 39: Conejos River near Mogote April 1st forecast percent error (for the 50% exceedance forecast of April-September volumes) vs. April-September precipitation anomalies for NRCS, as measured by the GMET forcing dataset.

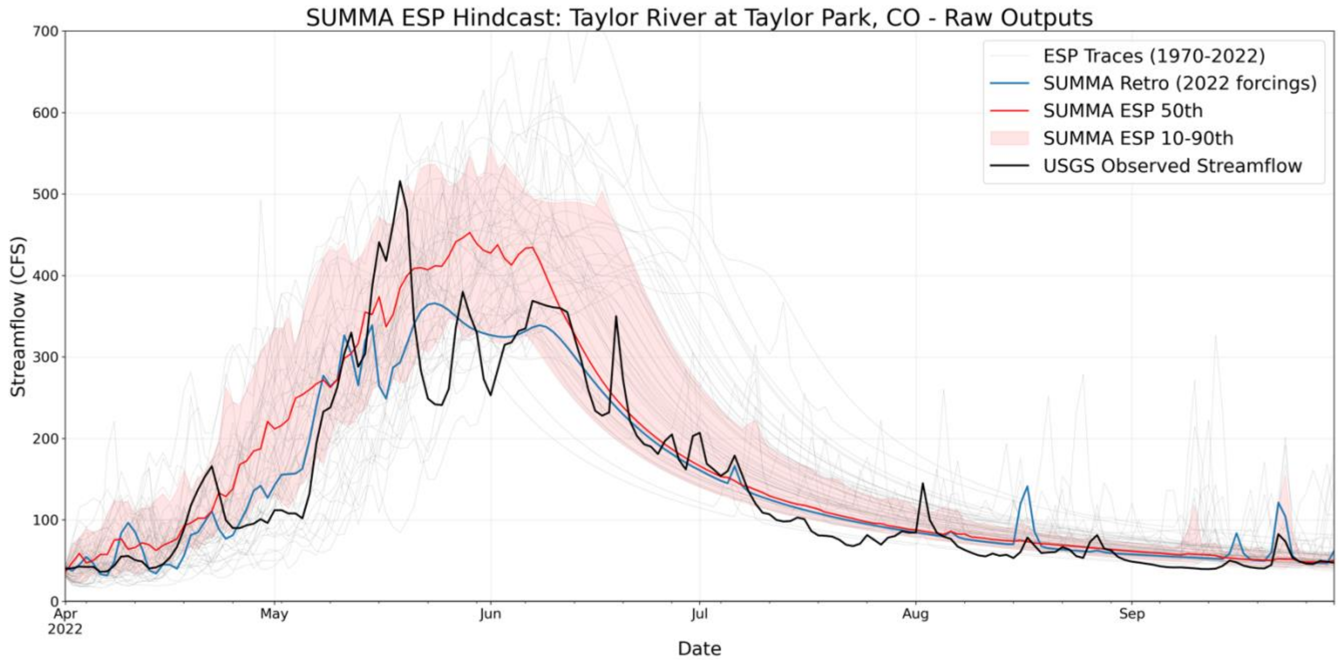
8.5. SUMMA Hindcasts

8.5.1. East River at Almont

Table 10: SUMMA simulation statistics for April 1st to September 30th, 2022, East River at Almont.

Water Year 2022 SUMMA Simulation Description	Nash Sutcliffe Efficiency	Kling-Gupta Efficiency
Retrospective (2022 forcings)	0.73	0.73
Retrospective (2022 forcings) w/ bias correction	0.92	0.88
ESP Hindcast median (2000-2021 forcings)	0.74	0.82
ESP Hindcast median (2000-2021 forcings) w/ bias correction	0.84	0.88

8.5.2. Taylor River at Taylor Park



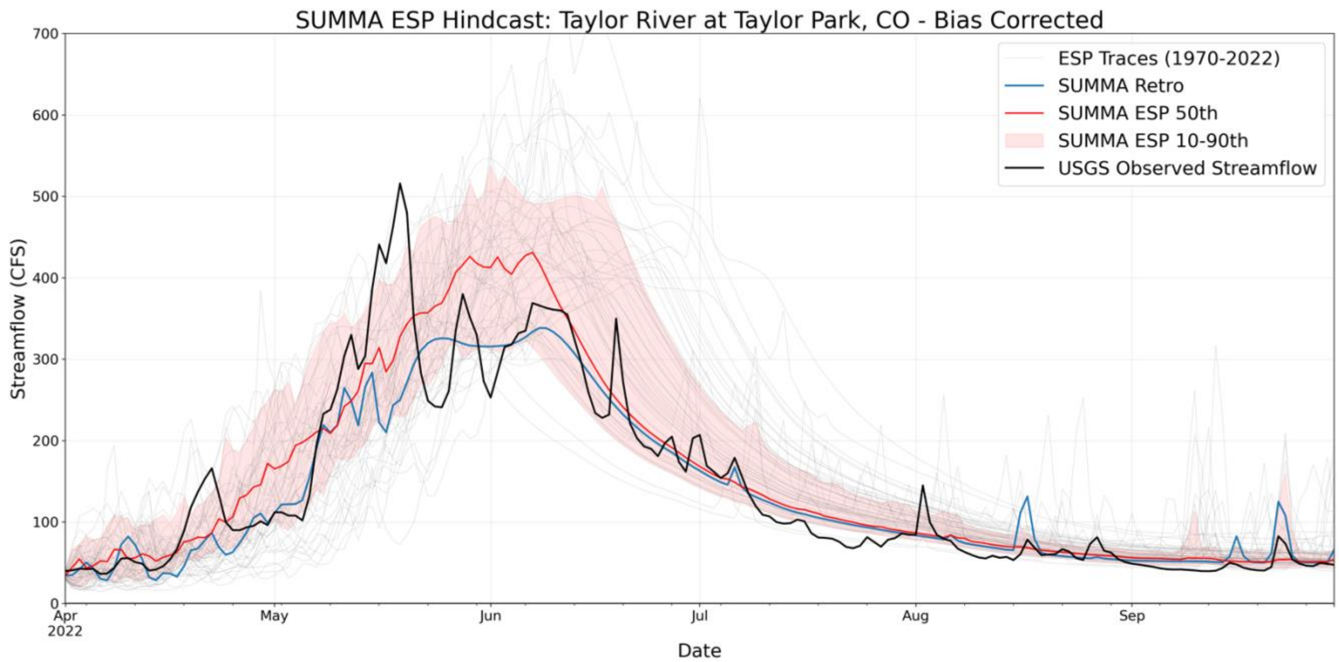


Figure 40: SUMMA raw (top) and bias-corrected (bottom) ESP hindcast of the Taylor River at Taylor Park for April 1st, 2022, showing 53 ensemble members from meteorological years 1970-2022 (grey traces) with respect to observed flows from the USGS gage (black). Probability distributions for meteorological years 2000-2021 are highlighted in red (10-90th percentile) where the median (50th percentile) is denoted by the solid red line. The 2022 retrospective model run (i.e., using actual 2022 forcings) is in blue. All simulations are bias-corrected using a Leave One-Out (LOO) approach.

Table 11: SUMMA simulation statistics for April 1st to September 30th, 2022, Taylor River at Taylor Park.

Water Year 2022 SUMMA Simulation Description	Nash Sutcliffe Efficiency	Kling-Gupta Efficiency
Retrospective (2022 forcings)	0.85	0.87
Retrospective (2022 forcings) w/ bias correction	0.83	0.81
ESP Hindcast median (2000-2021 forcings)	0.77	0.79
ESP Hindcast median (2000-2021 forcings) w/ bias correction	0.81	0.88

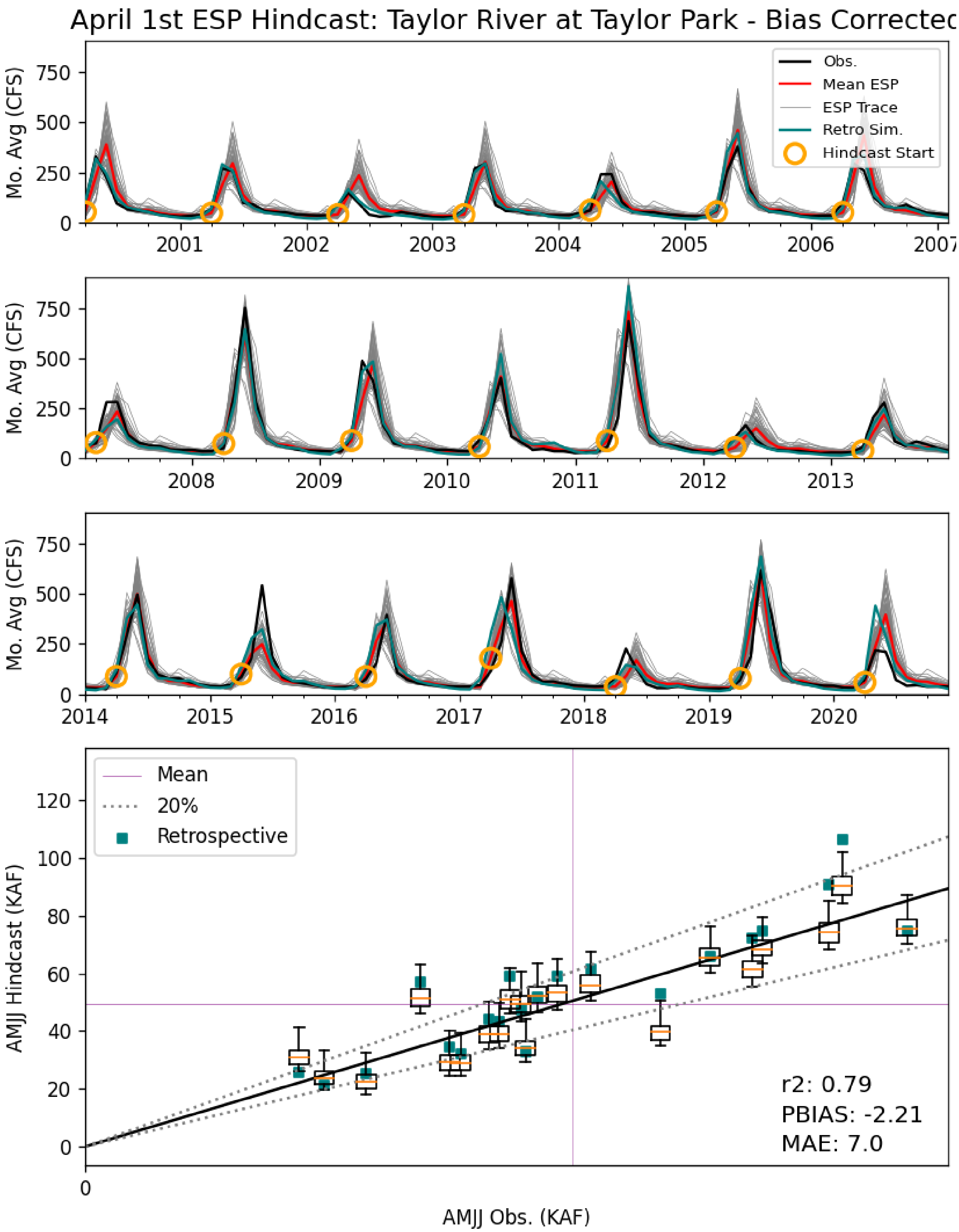


Figure 41: Multi-decadal SUMMA hindcasts (bias corrected-ESP for all April 1st dates) and retrospective simulations for the Taylor River at Taylor Park. Here, the top panels show the 2000-2022 time series; the bottom panel shows the volumetric April-July flows for the probabilistic hindcasts and deterministic retrospective simulation against USGS observations. Statistics are of the median BC-ESP hindcast.

8.6. CASM Streamflow Forecast Roundtable Agenda

CASM Streamflow Forecast Discussion Agenda – WY 2022

Nov. 7th, 2022, 3:00 – 5:00p MT

Meeting goals

To gather forecasters, modelers, and select forecast end-users to discuss the challenges and opportunities within seasonal streamflow forecasting in Colorado, specifically as it pertains to the integration of ASO snow data into hydrologic models.

Meeting outline

- Brief presentations and reflections on the 2022 forecasting season from official federal forecasters Gus Goodbody (NRCS) and Cody Moser (CBRFC) and WRF-Hydro modeler Dave Gochis (NCAR)
- Q&A with forecasters and modelers
- Open discussion

Broad questions

1. **What are some of the limitations and challenges with existing seasonal water supply forecast products? With integrating ASO data into these models?**
    - Official forecasts from federal agencies: NRCS, CBRFC, and WGRFC
    - Developmental or private forecasts: NCAR, water utilities, etc.

2. **What does an improved water supply forecast in Colorado look like?**
    - How much error is from snowpack vs. April-July precip anomalies, model calibration/structure, soil moisture, etc.?
        - CBRFC error attribution study

3. **What is a good baseline for adopting a new process operationally?**
    - Can we even evaluate a single forecast season?
    - How long might it take to get a reliably improved forecast?
    - What information do we need to begin to draw conclusions about experimental products?

8.7. CASM Streamflow Forecast Roundtable Minutes

CASM Streamflow Forecasting Roundtable
Meeting Notes – WY 2022

Nov. 7th, 2022, 3:00 – 5:00p MT

Organized by Lynker together with CASM Planning Team
Facilitated by Josh Sturtevant, Lynker Water Resources Scientist

Forecaster/Modeler Invited Presentations

1. **Angus Goodbody**: Water Year 2022 Streamflow Forecast Verification; USDA NRCS
    a. Last fall (2021)
        i. Soil moisture:
            1. Near four corners and a lot of Colorado = deficits in soil moisture
            2. Deficits in the San Juan Mountains – southside of the San Juans
        ii. Low median baseflows; SW Colorado, below to well-below flows
        iii. Precipitation (wet in Dec – in median – wet late spring into late summer) and SWE (fairly early/rapid melt)
    b. Verifications of April 1 forecasts
        i. April 1 forecasts were near median, below median
            1. Errors: over forecast in SW -> variable through Rio Grande -> variable in upper CO headwaters (impacts from fire – high runoff efficiency as a result – higher snowpack accumulation)
            2. Exceedance probability category: 50-70% bin in SW -> 70-90% (drier than what was forecasted) through Rio -> over forecast in Gunnison -> variable in headwaters
        ii. Precipitation (April – June period); near normal; above normal precip for portion of the Rio Grande
2. **Cody Moser**: CBRFC official and ASO-assimilated experimental forecasts
    a. Examples of experimental vs official SWE
        i. Dillon: experimental forecast on par (=) with official
        ii. Taylor: experimental < official
        iii. Dolores: Experimental > official
        iv. Pine: Experimental > official (ASO verifying better than model)
3. **Dave Gochis**: NRCS/ASO WRF-Hydro modeler
    a. East River, Upper Colorado Windy Gap: ASO can help; flow forecast is higher but upstream diversions pose problems with forecasting
    b. Taylor: within 3% of Apr-Sep value
    c. Forecasts generally stayed on target after May survey (Taylor, Conejos, Dolores, Blue)
        i. 50% of flow has already passed gage by then
    d. Active monsoon impacted basins
    e. Generally, among basins, it is a case-by-case impact of ASO assimilation
        i. For example, direct insertion did not work well in Conejos during second flight, due to mass balance issues. Monsoon helped forecast verify well, regardless.

- Forecasting snowpack -> translating that into a streamflow forecast? Perhaps this would be okay with an index-based approach, maybe targeted peak SWE ASO flights?

How can we better improve the forecasting component?

- Hindcast studies
  - Understand benefit of ASO data in model as compared to open loop (w/o ASO data)
  - Understand changes in predictability
- Error attribution studies (e.g., how much error actually came from spring precip [which ASO will not address] vs. from initial snow states in the model?)
  - CBRFC study suggests that in CO, 50% of Apr 1 forecast error is from spring precip
- Benchmark comparison of different models
- End of water year forecast verification efforts
  - Inter-agency/inter-group

How can CASM and CASM-supported streamflow forecasts better serve end users?

- What is your model calibrated to?
  - Clarity around calibration period/datasets (which streamflow data, where?)
    - Naturalized vs. observed streamflows
  - Issues in WRF-Hydro with naturalized forecasts without reservoir operations, which were not really useful to Northern Water, for example.
    - Need to set expectations, but also strive to better represent managed systems, where possible.
- If we account for just a few of the biggest diversions, then we can see where the model is working and where it is not, vs. just writing errors off to differences in naturalized vs obs flows
  - Most large diversions have real-time obs, so these can be factored into WRF-Hydro forecasts
  - Of course, observations also have some error– 5-8% uncertainty with USGS flows
  - Also, CBRFC spends lots of time building timeseries of adjusted flows, which WRF-Hydro could use for calibration purposes

- Also ASO critically needs a more seamless way to disseminate WRF-Hydro forecasts -> might be nice to not just share results in an email, but actually have a data portal- both for documentation, accessibility, and transparency- similar to how ASO SWE data are shared.
  - Strive to emulate the well-established protocols of CBRFC and NRCS
    - Easier to compare forecasts across models in the future when they are consistent across groups/models
    - For example, this year ASO WRF-H forecasts often times only included the 50th percentile, and not the full 10/30/50/70/90 distribution
  - Also, access to both open-loop and ASO DA WRF-Hydro forecasts is a goal
    - To ask: "What is the actual value (marginal benefit) of including ASO data in WRFH forecasts?"
    - Otherwise, hard to discern value of WRF-H from value of ASO data used in WRF-H

Concluding remarks/action items:

- Shared target volumes and periods between WRF-Hydro forecasts and official forecasts
  - Representation of major diversions needed in WRF-hydro
  - Similar forecast periods (e.g., AMJJ in most of CO, AMJJAS in Rio Grande)
- Develop data delivery protocols for WRF-Hydro
  - Not just emailed forecasts, but needs documentation/forecast data portal

Q&A

Challenges with direct insertion of ASO data into well calibrated hydro/snow model:

- (Cody Moser) How to calculate an ariel extent of ASO SWE? 3m gird? Set a threshold? As of now treating extent as the "best" way it would occur
- What we had thought about the model "getting true snowpack" … sometimes it is, sometimes it isn't. Do not know bias yet to obtain bias adjustment.
    - *Unfettered direct insertion does not look like the right thing to do – as much as we craved to do it!*
    - *Some instances where it is okay, e.g., Big Piney WY (under-gaged basin), but overall, it gives inconclusive results at best.*

Data assimilation techniques that WRF-Hydro team is considering (due to direct insertion limitations):

- (Dave Gochis) Tradeoffs by using direct insertion; pros/cons of modelling
    - Con of direct insertion: Lack of mass continuity (e.g., Conejos in May WRFH forecast)
    - Particle filter methods can get you "closer" BUT computational expense is a con
    - Simplifying the model lessens computational issues
        - WRF-Hydro is inherently computationally expensive to run (at 1km resolution)
- More frequent flights may get around this– data assimilation windows are shorter – but then you end up paying too much for undetermined marginal benefit.

What is a good baseline for adopting these new models and processes operationally? What does the roadmap look like to get to that point?

- Susan Behery (Reclamation)
    - Uses ASO data but alongside existing forecast data: *Why? Because* ***time and comfort of official products, esp CBRFC***: Understanding the nuisance year to year
        - What do typical forecasts look like? What is the added value? How much closer is it to the **existing** forecast (very refined forecast) that Reclamation uses? What is the **ease** of use?
        - When there is an agreement – there is increased confidence!
- Russell Slade (Denver Water) uses ASO data to validate forecasts
    - Take in as much information from ASO data on daily basis for tracking and monthly basis for operational

NRCS – statistical model – how can ASO data be incorporated as a SWE predictor into statistical models that require long periods of record in the predictor variables?

- Use max 30 years of data for model calibration, traditionally – now use at least 10 years of data (pragmatically determined)
    - *Need to have a balance of # of years for basic requirements*
    - ASO data has a short period of record so it is hard to assimilate into model, without using an indexing method
- Sean Flemming: Find a way to get ASO data incorporated into new NRCS machine learning forecast framework
    - Reconstruct ASO-like product on the basis of relationships over the historical data and then train the existing model; compare that with the SNOTEL data

What kind of reliability are we looking for (ASO flights, vs. building a station-based index to historical ASO flights)?

- Building a period of record – calibrating to that… make that the new calibration normal? **Based on assumption of stationarity**
    - Spatial patterns can change! See: California last year where majority of snowfall was from one AR early in the winter
- To what accuracy are we pushing for?