Towards Improved Verification and Metrics: A Case Study Over the Sevier River Basin at the CBRFC

Camden Opfer, 2023-25 Ernest F. Hollings Scholar

Co-Mentors Paul Miller and Michelle Stokes

National Weather Service; Colorado Basin River Forecast Center

Weather-Ready Nation

July 24, 2024

Introduction

The Colorado River Basin Forecast Center (CBRFC) is charged with predicting both the near-future flow, annual peak flow, and total water supply from April to July for sub-basins across the Colorado River, Great, and Sevier River Basins, informing policy makers, local officials, dam operators, and other stakeholders throughout the region. This impact spans from my hometown of Boulder, Colorado to the location of the CBRFC in Salt Lake City, Utah, and to the many farms, ranches, and vineyards of Southern California.

Man-made dams, diversions, canals, reservoirs, and other structures within these basins form a vast, interconnected network which allows for aggressive water management. This infrastructure facilitates resiliency and preparedness for day-to-day, seasonal, and interannual changes in meteorological and hydrological conditions, whether there is an incoming storm system with flood potential or a decades-long drought. The effectiveness of this infrastructure is determined by the skill of planners, managers, and forecasters who decide how to apply it. As a leading forecaster for the region, the CBRFC plays a large role in the weather preparedness of these basins, and improvements to the CBRFC's forecasts have a direct impact by improving the data used as a basis for planning and response efforts.

Though hydrologist and meteorologist calibration and interpretation play a role in creating the official forecast values published by the CBRFC, computational modeling is key to resolving the trends and bounds of water supply predictions. The primary model used at the CBRFC is the Sacramento Soil Moisture Accounting (SAC-SMA) model, which numerically predicts surface processes ranging from snowmelt to subsurface storage to evapotranspiration (Burnash, 1973). The most impactful meteorological inputs to hydrologic models like SAC-SMA are temperature and precipitation, as they determine the snowpack accumulation over the winter, the melt of snowpack through the spring and summer, the additional moisture brought in by the summer monsoon, and the evaporation of surface water throughout the year (CBRFC, 2020). To capture the interannual variability of temperature and precipitation, the Ensemble Streamflow Prediction (ESP) configuration is used, in which the SAC-SMA model is fed the historical meteorological conditions of each year from 1991 to 2020, which each correspond to an ensemble member. This results in a 30 member ensemble with meteorological forcing ranging from extremely dry to extremely wet years. Figure 1 shows an example of the seasonal progression of volume forecast by each of these members.

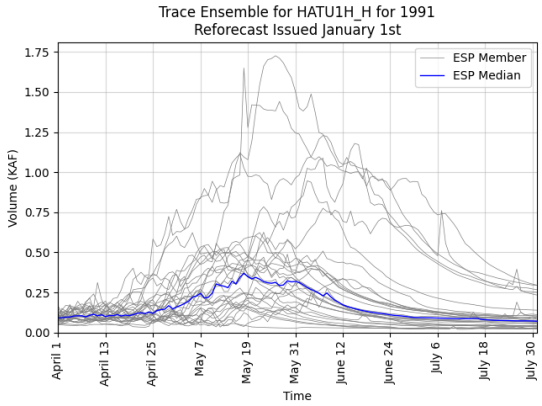


Figure 1: Ensemble traces for a regulated reforecast of 1991 April to July volume for the Sevier River near Hatch, UT

The forecasts produced by the CBRFC are generally unregulated, meaning they represent what the total volume would be, while removing the effects of measured upstream human interventions. However, the forecasts for some locations within the Sevier River Basin in Southern Utah are regulated, attempting to match the value observed in the real world by accounting for both measured and unmeasured upstream intervention. This unique modeling approach is applied because of the unique characteristics of the basin. These include its low population of around 80,000 people (US Census Bureau, 2020), the dominance of agricultural irrigation as a water use, and the complete exhaustion of water resources to meet these user demands, leaving negligible amounts of water to reach the terminal Sevier Lake except during extremely wet years (Williams et al., 2014). While residential use is thoroughly measured by utilities providers, smaller agriculture-focused diversions often have less formal controls, may not have real-time flow measurement, and a meaningful amount of the water diverted is returned as runoff. This is particularly relevant in the Sevier River Basin, because the low overall supply and small residential demand mean that unmeasured diversions make up an unusually large share of claims. As a result, the observed unregulated volume can be expected to differ significantly from a regulated amount, and the prominence of unmeasured interventions in particular provides a potential source of uncertainty and error.

Hypothesis

This investigation seeks to answer the question: Are regulated or unregulated forecasts more skillful at predicting water supply in the Sevier River Basin? Intuitively, it can be expected that the unregulated forecasts are more skillful, since these forecasts avoid the added uncertainty of human action.

Methods

Data

Historical observed flows come from stream flow gages operated by the U.S. Geological Survey, which make their data available freely online (https://dashboard.waterdata.usgs.gov/app/nwd/en/) and through an API (https://waterservices.usgs.gov/docs/dv-service/daily-values-service-details/) for easy access (U.S. Geological Survey, 2024). The total flow accumulated from April through July of each year is the main value that the CBRFC produces supply forecasts for, so a sum over time was used to match this. For this investigation, the USGS data was retrieved from the CBRFC's internal database, and calculations to extract unregulated observations were carried out. Because of data availability for unmeasured diversions, high-quality observations were limited to 1991-2015.

Both unregulated and regulated ESP reforecasts were made by hydrologists at the CBRFC for the purpose of this project. This includes 30 ensemble members with meteorological conditions from 1991-2020, when high-quality meteorological data is available, each used to make a forecast for the years 1991-2015. This resulted in one .csv file per location and forecast issuance date and, similar to the observational data, these were summed across the April through July forecasting period for the purpose of this investigation. This resulted in one value per ensemble member per forecast issuance date at each location.

All of this data was acquired for the HATU1, SEKU1, and SRKU1 gage locations, which measure three of the four sub-basins of the Sevier where the CBRFC currently issues regulated forecasts. The fourth, PIUU1, had insufficient data to be included in this investigation, and is located at a reservoir just downstream of SEKU1 and SRKU1.

These observations were supplemented with information from the Sevier Water Users Association, which provides more localized reporting of diversion and use volumes (SRWUA, 2024).

Verification Metrics

The CBRFC official website (https://www.cbrfc.noaa.gov/) contains several pages dedicated to model verification— comparing a reforecast of water supply to the observed volume for past years. This includes a spatial map showing the mean absolute percent error (MAPE) at each forecast location, as well as plots specific to the locations. See Figure 2 for a sample of these verification plots, which are created from the official, single value forecast and the percentile ranges of the ESP reforecast ensemble at each location.

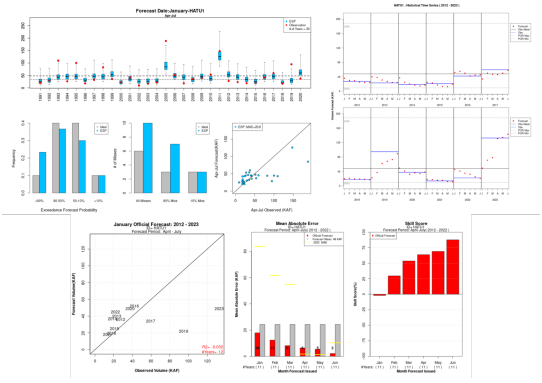


Figure 2: Verification plots on the CBRFC official website for the Sevier River near Hatch, UT

The general goal of these metrics is to show the skill of the ensemble at predicting the total flow at a particular location from April through July. Forecasts for this period are issued on the first day of each month from January through June, and the distinct meteorological forcing of the 30 years gets replaced by true observations for the particular year as time progresses and

they become available. Because of this, modeled conditions approach truth for that year as the months progress, which is why the forecast issued in June can be expected to be much more accurate than the one in January.

Over the course of this investigation, a Python library containing code to recreate the CBRFC's existing plots, as well as various new verification plots and metrics, was developed. Applying this work, over 90 figures can be made for each location, though some have been found to be more helpful than others. As a result, not all of these new plots are described here, with a focus instead placed on the broader statistical approaches used to create them.

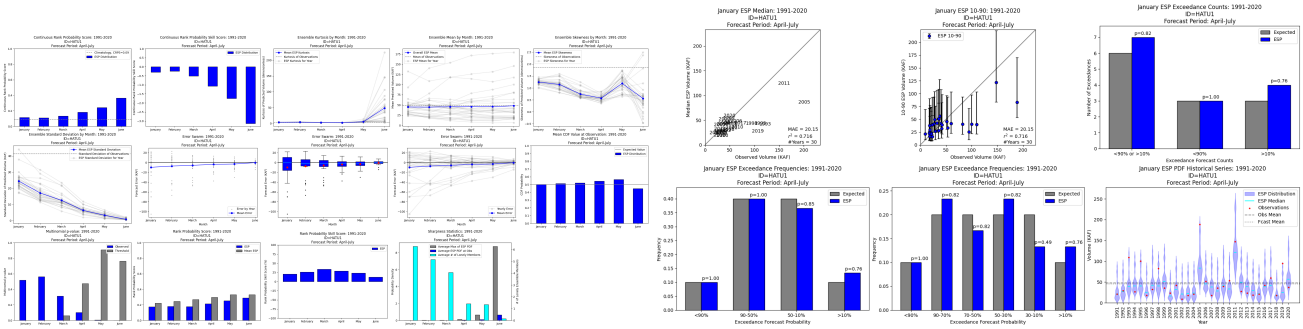


Figure 3: Newly devised plots for verification, created for the Sevier River near Hatch, UT

One new methodology creates and analyzes an ensemble probability density function (PDF), rather than the 10th, 30th, 50th, 70th, and 90th percentile values which were the focus of previous verification efforts. Following Brocker and Smith, 2008, each ensemble member is used as the mean of a normal distribution, called a kernel, with standard deviation σ_k, which is often called bandwidth in this application. The PDF for the ensemble at a given point x is then the mean of all the kernels' PDFs evaluated at x. This is illustrated in Figure 4, which takes the 30 discrete values of the ensemble, shows the PDFs for the individual kernels, and finally shows the overall, interpolated PDF.

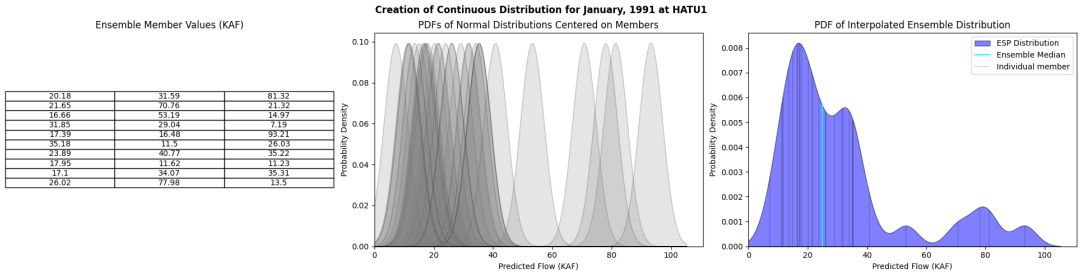


Figure 4: A plot demonstrating the creation of a continuous distribution from a discrete ensemble using regulated forecast data for January, 1991 at the Sevier River near Hatch, UT

This interpolated distribution will have the same mean as the mean of the discrete ensemble members, since the probability density above and below each member is symmetric. The variance of the new distribution is slightly more complicated to derive, but Brocker and Smith, 2008 provides the general value of $\sigma^2 = \sigma_k^2 + Var(X)$, where $Var(X)$ is the sample variance of the set X, containing the discrete predictions by the 30 ensemble members. This value is always greater than the variance of the underlying ensemble because interpolation widens each member from a single point to a full normal distribution. In the implementation for this investigation, bandwidth is chosen to be $\sigma_k^2 = \frac{Var(X)}{|X|}$, where $|X|$ is the cardinality of X. This yields an overall variance which is a function of $Var(X)$, given by

$$\sigma^2 = \sigma_k^2 + Var(X) = \frac{Var(X)}{|X|} + Var(X) = \frac{Var(X)(1+|X|)}{|X|}.$$

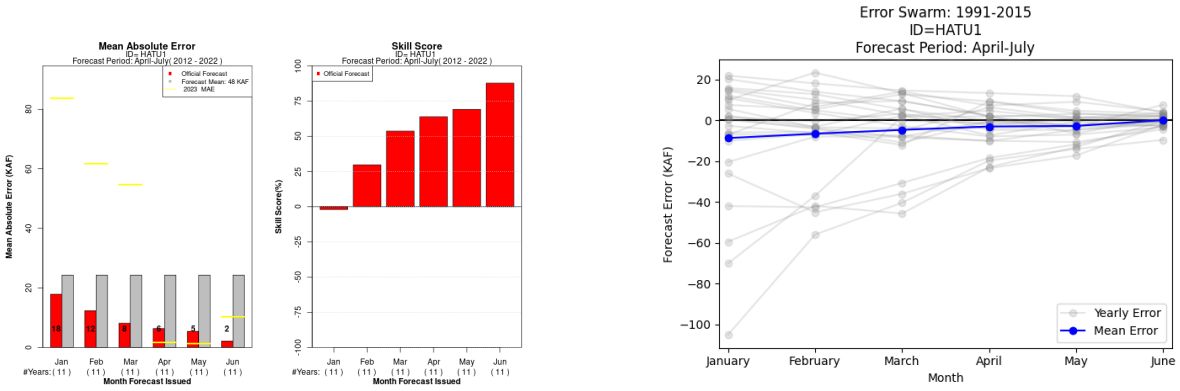
Computing this interpolated distribution allows a new subset of plots and statistics to be generated. One of these statistics is continuous ranked probability score (CRPS), which is found with

$$CRPS = \overline{\int_{-\infty}^{\infty} \left(F(x) - 1_{\{x>obs\}}(x)\right)^2 dx}$$

where $F(x)$ is the cumulative density function (CDF) of the ensemble and $1_{\{x>obs\}}(x)$ is an indicator function which is 1 when x is greater than the observed value and otherwise 0. The overline here represents taking the mean across all years of forecast data, or whatever other spatial or temporal domain is suitable.

Some intuition for this statistic is that the ideal CDF of a forecast would start at 0, then instantaneously jump to 1 at the x value of the observation, indicating 100% certainty in that value occurring. This will never happen for a real-world ensemble forecast, so the CRPS sums the difference between the model's CDF and the ideal one. Therefore, an ideal model would have a CRPS of 0, while a highly unskilled model would have a CRPS approaching 1. Though the CRPS is just one number, it is a well-rounded representation of skill because it rewards models with a correct (or nearly correct) median value, while penalizing models which over- or under-predict extremes, something which is particularly relevant in water management, where extremes of drought and flood have severe impacts. This makes the interpolation approach quite useful, as it allows for more holistic assessment of all the ensemble's members.

Another new approach has been creating swarm plots which show how one statistic progresses by forecast issuance month for each year, in addition to the averaged values across years. For example, a bar chart showing mean absolute error and mean absolute error skill score has been produced in the past, focusing on values averaged across years. This is shown as Figure 5a, while a new plot with mean error averaged across years, is shown as Figure 5b. By including both the average value and individual contributions of each year, a better understanding can be reached of whether error is determined mostly by outlier (extremely wet or dry) years, and how well the ensemble converges through the months.



(a) Existing MAE Plot

(b) New Error Swarm Plot

Figure 5: Comparison between existing and new error visualization plots

Similar plots have also been made for the mean, standard deviation, skewness, and kurtosis of the interpolated distribution, giving insight into the progression of the model's best guess, spread, and deviation from a Gaussian distribution. In all, these new methods aim to provide more ways to evaluate the skill of ESP forecasts and reforecasts, or to distinguish between two modeling approaches. The latter is carried out here by producing equivalent plots and statistics for both the regulated and unregulated reforecasts, then comparing their skill.

Results

Sevier River at Hatch, Utah (HATU1)

Near its headwaters, the Sevier River passes by the town of Hatch, which has a USGS gage continuously measuring the river's flow. Because of the mountainous terrain, little water is diverted prior to reaching this gage, since there is limited human activity in the area. This can be seen in Figure 6, which indicates the total volume of estimated water taken by diversions is 8.30% of the observed, unregulated volume. This indicates that diversions play a limited role in this sub-basin, and that regulated and unregulated volumes will be similar, which is also visualized in Figure 6, where the blue and black lines are almost identical. Note that the unit KAF refers to thousands of acre feet, a measure of volume where one acre foot is the volume of one acre of area covered uniformly by 1 foot of water, which is 43560ft^3 or 1233.5m^3.

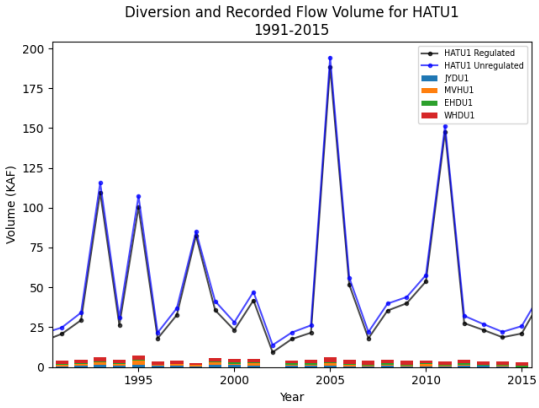
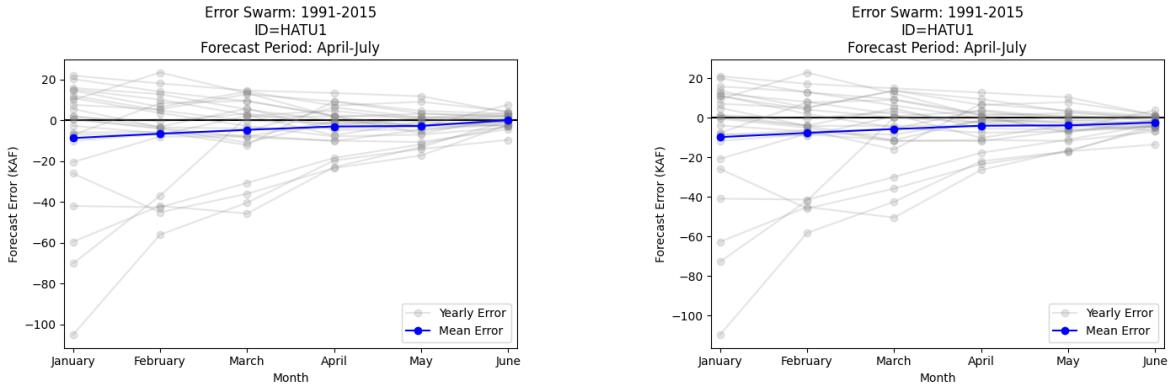


Figure 6: The magnitude of diversions, regulated flow, and unregulated flow by year for the Sevier River at Hatch, UT

Because so few diversions differentiate regulated and unregulated forecasts of the HATU1 volume, there is minimal difference in forecast skill between the two forecast types. This can be seen in Figure 7, which shows equivalent trends in mean error (bias) of the model as the months progress, and very similar results for each particular year. The tables in the All Locations section will show that the two forecast types are also similar in other metrics. Overall, the similarities in regulated and unregulated forecasts at HATU1 mean that this site should not play a large role in the choice of operational forecast type.



(a) Regulated

(b) Unregulated

Figure 7: Median ensemble member error by month for the Sevier River at Hatch, UT

Main Stem of the Sevier River at Kingston, Utah (SEKU1)

Unlike the HATU1 station, the sub-basin volume measured at SEKU1 has significant diversions. Figure 8 visualizes this, with the bars representing diversions reaching closer to the observed flow at the gage, sometimes having greater magnitude than the regulated volume. Here, the volume diverted is, on average, 60.% of the unregulated volume for the year, and is occasionally greater in magnitude than the regulated volume. This results in much more differentiation between regulated and unregulated observations, something which the two forecast types aim to match.

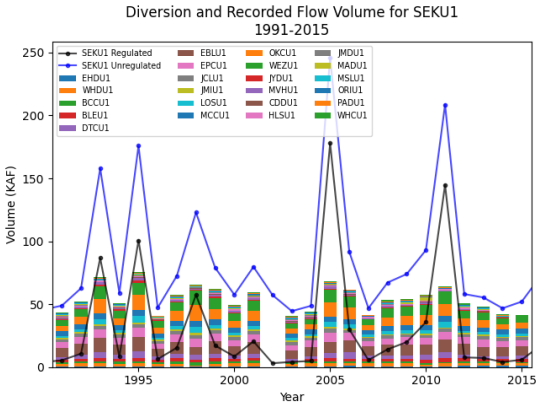
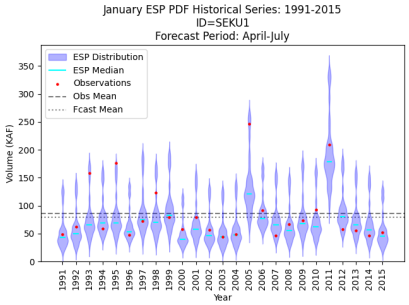
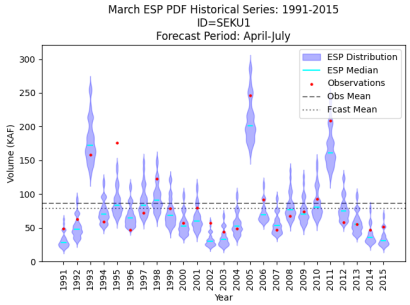


Figure 8: The magnitude of diversions, regulated flow, and unregulated flow by year for the Main Stem of the Sevier River at Kingston, UT

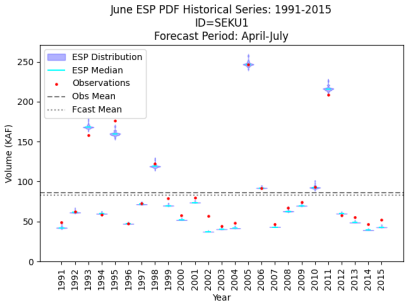
Figure 9 is a visual representation of the ensemble's PDF by year, with thicker regions of the violin plot corresponding to greater probability density for the ensemble. So, the ideal plot would have one particularly wide bump at the value of the observed volume, which is represented by the red dots. Looking at these particular results, the unregulated observations do have a higher mean and median value than their unregulated counterparts, and this trend is exacerbated in the extremely wet years of 1995, 2005, and 2011. So, though the standard deviation of the two model types approaches the same value as the months progress, this value is a smaller percentage of the unregulated flow, leading to overconfident results for June reforecasts.



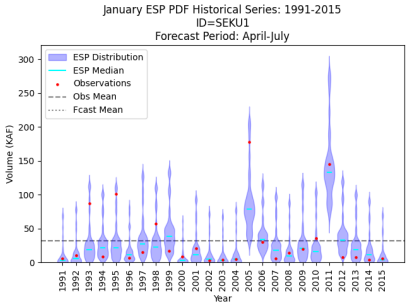
(a) Unregulated January



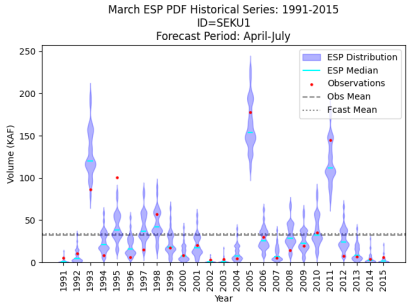
(b) Unregulated March



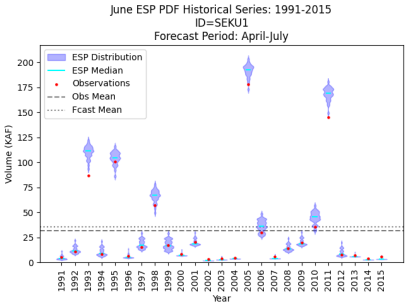
(c) Unregulated June



(d) Regulated January



(e) Regulated March



(f) Regulated June

Figure 9: Violin plots of the interpolated distributions derived from reforecasts issued in January, March, and June

Though June unregulated forecasts struggle with producing realistic spread, the accuracy of both model types is good. All January forecasts contain the observed value within the bounds of the distribution, and the only March forecast which doesn't is 1995, a particularly wet year. In June, both reforecasts have a slight but consistent low bias, though this is more significant for unregulated forecasts, which have an average CDF above 0.7 at the value of the observations. This means that 70% of the area of the violin plot is below the true volume for the average year. It is difficult to say if manual forecaster correction of ensemble spread would rectify this issue, but regardless, both regulated and unregulated forecasts show promise for this location.

Eastern Fork of the Sevier River at Kingston, Utah (SRKU1)

Though not as dominant as the diversions upstream of SEKU1, diversions before SRKU1 play a large role in its hydrology. On average, the volume diverted is 43% of the unregulated volume for the year, and reaches as high as 60.% for 1996. So, similar to SEKU1, a significant difference between regulated and unregulated diversions can be expected.

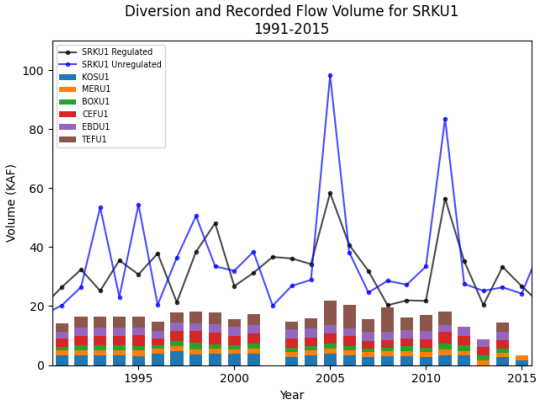
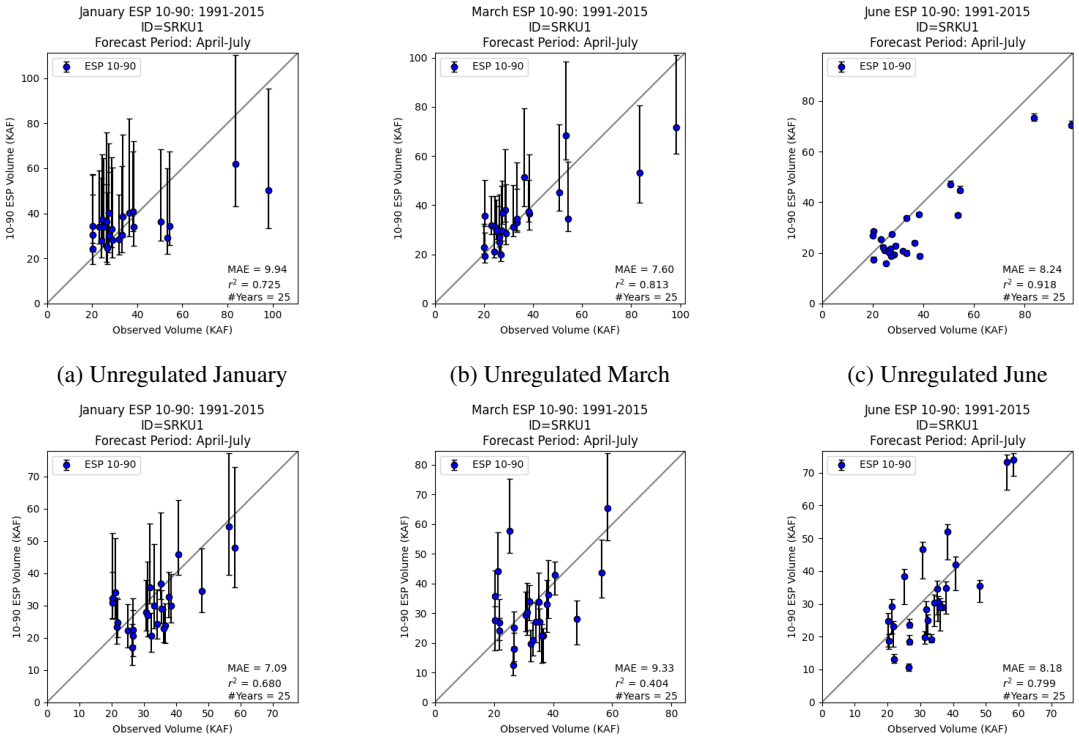


Figure 10: The magnitude of diversions, regulated flow, and unregulated flow by year for the Eastern Fork of the Sevier River at Kingston, UT



(a) Unregulated January (b) Unregulated March (c) Unregulated June

(d) Regulated January (e) Regulated March (f) Regulated June

Figure 11: Correlation between forecast and observations plots for reforecasts issued in January, March, and June

Figure 11 conveys some of the most significant changes between regulated and unregulated observations and forecasts at this point. The x-axis is observed volume which, for unregulated observations, is mostly confined to 20-40KAF, with outliers near 100KAF. On the contrary, regulated volume is also predominantly in the 20-40KAF range, but peaks just above 60KAF. This disparity comes from greater unmeasured diversions in particularly wet years, motivated perhaps by management to reduce high flow and flood risk, or by a rotation towards more irrigation dependent crops.

On the y-axis of this figure is the reforecast value, with error bars used to show the 10th and 90th percentile envelope of the ensemble. This 10-90 range is a standard in river forecasting, representing some of the most extreme values that most water planning should readily account for. For the purpose of model verification, these 10-90 ranges should intercept the true, observed values (represented by the line $y = x$) around 80% of the time, which is the case for both forecasting types in January and March. However, in June, the ensemble distributions for both regulated and unregulated forecasts have standard deviations averaging 2KAF, which manifests itself in Figure 11 as overly tight 10-90 ranges. This is because the spread of the ensemble is based solely on meteorological conditions, which are consistent across years since it will almost certainly be hot and dry from June through the end of the forecast. This neglects uncertainty in the state of each sub-basin's snowpack, tension water, free water storage, and runoff capacity, which are never directly measured. This type of error is accounted for manually when hydrologists issue official forecasts, relying on field knowledge and historical outcomes to determine reasonable values.

Overall, though the unregulated reforecasts are better correlated with observations in each issuance month, with higher r^2 values throughout, they are not as representative as regulated reforecasts, which have more realistic spread. This issue would be compensated for in an operational setting, but may skew the summarizing statistics described in the following section.

All Locations

By summarizing across years, it is more manageable to compare the skill of regulated and unregulated forecasts between locations and issuance months, producing more holistic evaluation of forecast accuracy. To this end, tables of mean absolute error (Table 1), mean absolute relative error (Table 2) and continuous ranked probability score (Table 3) have been constructed. Since smaller values of these statistics are favorable, the lower of the two forecast types have been highlighted in each column, showing which has a better value for a particular year and location, and overall through averaging. An average without June is also provided, since those forecasts sometimes have quite different characteristics than the other issuance months.

The first of these tables shows mean absolute error (MAE) values, calculated as $MAE = \overline{|obs - rfcst|}$, with observed volume obs and median ensemble member volume $rfcst$. As a result, the ideal value would be 0, representing $obs = rfcst$, while greater values show a greater magnitude of error. Now focusing on specific values, HATU1 shows slightly better MAE for all months, though it remains a tossup as can be expected given the small number of diversions creating little distinction between regulated and unregulated reforecasts. The SEKU1 station has a much stronger bias towards the regulated forecasts in all months, with the regulated reforecasts having only one-sixth the MAE value of the unregulated reforecasts for June. Lastly, the SRKU1 station has MAE values which, overall, slightly favor the unregulated forecast. Altogether, these results combine to indicate that MAE supports the use of regulated forecasts, though this conclusion is not unanimous or overwhelmingly strong across space and time.

Issuance Month	Forecast Type	Jan	Feb	Mar	Apr	May	Jun	Avg	Avg Without Jun
HATU1	Unregulated	20.11	15.46	12.90	7.79	6.16	3.15	10.93	12.48
	Regulated	19.77	15.14	12.25	7.67	5.67	2.61	10.52	12.10
SEKU1	Unregulated	25.37	21.11	19.47	13.5	17.18	30.48	21.18	19.33
	Regulated	18.22	13.04	11.48	8.55	6.14	4.65	10.35	11.49
SRKU1	Unregulated	9.94	8.92	7.60	6.38	6.99	8.24	8.01	7.96
	Regulated	7.09	9.25	9.33	9.91	9.31	8.18	8.85	8.98
Average	Unregulated	18.47	15.16	13.32	9.22	10.11	13.96	13.37	13.26
	Regulated	15.03	12.48	11.02	8.71	7.04	5.15	9.90	10.86

Table 1: Mean absolute error by location and forecast issuance date

Though mean absolute relative error (MARE) is similar in calculation to MAE, the normalization by annual observed flow mean that a large magnitude of absolute error can be acceptable in particularly wet conditions, while less error is tolerated in drier years. This normalization can be expected to provide more representative statistics than MAE since, as described previously, unregulated observations have much wetter extremes, which would allow for a much greater magnitude of error for the same relative error. This results in quite different results in Table 2, where almost every month at all three locations favors the unregulated reforecast. The only exceptions to this rule are January at SRKU1, which weakly favors the regulated forecast, and June at the HATU1 and SEKU1. It is difficult to verify what caused the June reversal, but it is likely related to the overly tight ensemble or the use of observations to backfill April and May volumes (before the forecast issuance date), which may be inconsistent with the model's expected states of properties like free water storage and snowpack. Overall, though, MARE strongly favors the unregulated forecast across space and time, and since the normalization it applies makes for fairer comparison between regulated and unregulated reforecasts, this result should be given more weight than the MAE values.

Issuance Month	Forecast Type	Jan	Feb	Mar	Apr	May	Jun	Avg	Avg Without Jun
HATU1	Unregulated	0.357	0.278	0.262	0.170	0.144	0.084	0.216	0.242
	Regulated	0.420	0.321	0.283	0.197	0.159	0.077	0.243	0.276
SEKU1	Unregulated	0.235	0.219	0.223	0.169	0.249	0.447	0.257	0.219
	Regulated	0.822	0.740	0.614	0.535	0.426	0.190	0.555	0.628
SRKU1	Unregulated	0.264	0.240	0.198	0.176	0.189	0.230	0.216	0.213
	Regulated	0.232	0.311	0.316	0.316	0.302	0.253	0.288	0.295
Average	Unregulated	0.286	0.245	0.228	0.172	0.194	0.254	0.230	0.225
	Regulated	0.491	0.457	0.404	0.349	0.296	0.173	0.362	0.400

Table 2: Mean absolute relative error by location and forecast issuance date

The final table produced for all locations shows the continuous ranked probability score (CRPS), the calculation of which is described in the Verification Metrics section. Like MAE, this statistic is not normalized by the value of observations. However, rather than find only the distance between the median of the ensemble and the observed value, CRPS characterizes the error of the ensemble as a whole. This means that normalization is much less important. Further, the process of normalization adds another layer of abstraction as well so, for a statistic like CRPS that is already quite abstract, it is arguably more valuable to have the directly computed values. Thus, while a normalized form of CRPS has also been generated, it is not shown here.

Much like with MAE, the results in Table 3 show that the choice between regulated and unregulated forecasts at HATU1 is essentially a tossup, while SEKU1 has a stronger favoring of regulated forecasts. Even stronger, though, is the favorability of unregulated forecasts at SRKU1, so a holistic assessment may slightly favor unregulated forecasts based on CRPS. If the uncertainty added into official forecasts by hand could be replicated systematically, one would expect the wider spread of June unregulated forecasts to make this a stronger conclusion. For the purpose of this work, though, MARE provides much stronger conclusions than what CRPS yields.

Issuance Month	Forecast Type	Jan	Feb	Mar	Apr	May	Jun	Avg	Avg Without Jun
HATU1	Unregulated	0.106	0.101	0.133	0.178	0.262	0.394	0.196	0.156
	Regulated	0.107	0.100	0.130	0.176	0.240	0.343	0.183	0.151
SEKU1	Unregulated	0.097	0.089	0.136	0.204	0.269	0.503	0.216	0.159
	Regulated	0.091	0.066	0.111	0.172	0.254	0.253	0.158	0.139
SRKU1	Unregulated	0.087	0.079	0.108	0.161	0.174	0.538	0.191	0.122
	Regulated	0.097	0.148	0.215	0.237	0.275	0.439	0.235	0.194
Average	Unregulated	0.097	0.090	0.126	0.181	0.235	0.479	0.201	0.146
	Regulated	0.098	0.105	0.152	0.195	0.256	0.345	0.192	0.161

Table 3: Continuous ranked probability score by location and forecast issuance date

Between these three tables, it has been shown that regulated forecasts have smaller magnitude of error, but that unregulated forecasts have preferable error as a percentage of observed volume and a more ideal ensemble distribution as a whole. These results are encouraging, indicating that unregulated forecasts can provide more skill in this region.

Conclusions

To summarize, it has been found that regulated reforecasts are effective at reducing absolute error, particularly at SEKU1, while unregulated forecasts result in greater correlation with observations and less relative error, especially for SRKU1. Unregulated reforecasts have the outstanding issue of converging too aggressively in June, but this would be addressed by hydrologists issuing forecasts if they were to use unregulated ESP predictions as the basis of their forecasts. Unregulated and regulated forecasts are both usable by water managers, so the increased quality of forecasts when using an unregulated mode can be expected to improve decision making capabilities.

One weakness in this conclusion is that only 25 years of data at three locations is being applied, leaving uncertainty in how well the two forecast types will compare as climatology continues to shift, and water demands continue to rise. Despite the somewhat low spatial and temporal range of data accessed, robust inferences have been made by being sure to encapsulate the value of each of the 30 ensemble members through the use of CRPS and other metrics reliant on an interpolated distribution.

A more practical obstacle is the treatment of data from diversions with historical, as opposed to realtime, measurements. Many diversions have gages or other reporting mechanisms which give the CBRFC and other forecasters instantaneous information about flow, which can be used to improve forecasts looking ahead into the year. However, some diversions require calculations at the end of the runoff season to know what volume was taken. While both types of diversions were used in the differentiation of unregulated and regulated reforecasts here, a forward-looking unregulated forecast would not have access to this second type of data, leading to somewhat different results. This issue is an open question, potentially resolved through collaboration with the Sevier River Water Users Association, installation of gages, or estimation of flow based on mean historical values.

Another progression for this work would be to apply a similar methodology to the Colorado River and Great Basins also served by the CBRFC. This expansion would be valuable because these basins cover much larger geographical regions, with the Colorado providing 36 million people and 5.5 million acres of agricultural land with water (Bureau of Reclamation, 2013). Further, the Sevier River Basin is somewhat unique in that it's an isolated, terminal basin with low overall water supply and high agricultural demands. So, while the results found here could be expected to transfer to other basins, sweeping conclusions cannot be made without further study.

Future research could also include surveying users of CBRFC products and website visitors about which verification plots they find most relevant, useful, and readable. Along with this, seeing what conclusions these users draw from any given plot could provide insight into which verification metrics and figures are most effective at communicating the skill of the SAC-SMA model and the distribution of the ensemble. Overall, this work would improve transparency and help decision makers understand the level of certainty associated with future predictions.

In all, this investigation has shown that a transition to unregulated forecasts would improve accuracy and reliability, better informing water users, planners, and decision makers to appropriately respond to the unique hydrological and meteorological conditions of each year.

References

Brocker, J., & Smith, L. A. (2008). From ensemble forecasts to predictive distribution functions. *Tellus A*, *60*(4), 663–678. https://www.tandfonline.com/doi/abs/10.1111/j.1600-0870.2007.00333.x

Bureau of Reclamation. (2013, October). *Quality of water: Colorado river basin* (Progress Report No. 24). Bureau or Reclamation. https://www.usbr.gov/uc/progact/salinity/pdfs/PR24final.pdf

Burnash, R. J. (1973). *A generalized streamflow simulation system: Conceptual modeling for digital computers*. US Department of Commerce, National Weather Service; State of California, Department of Water Resources.

Colorado Basin River Forecast Center. (2020, October). *Model sensitivity analysis: An overview of cbrfc's hydrologic model sensitivity to changes in precipitation, temperature, soil moisture, and evapotranspiration perturbations* (Project Report). National Weather Service. https://www.cbrfc.noaa.gov/report/CBRFC_Model_Sensitivity_Analysis_2020.pdf

Sevier River Water Users Association. (2024). *Station tabular reports*. http://www.sevierriver.org/reports/station-tabular-reports/

US Census Bureau. (2020). *2020 census demographic profile*. United States Census Bureau. https://www.census.gov/data/tables/2023/dec/2020-census-demographic-profile.html

U.S. Geological Survey. (2024). *Water services: Statistics service*. United States Geological Survey. https://waterservices.usgs.gov/docs/statistics/statistics-details/

Williams, F., Chronic, L., & Chronic, H. (2014, April). *Roadside geology of utah* (2nd ed.). Mountain Press Publishing Company.