JULY 2023

Independent Peer Review of the State Water Project – Delivery Capability Report

An individual letter review for the Delta Science Program

Prepared by

Jon Herman – University of California, Davis

Delta Science Program

DELTA STEWARDSHIP COUNCIL

Table of Contents

Summary

This report develops a statistical method to adjust historical hydrology to account for recently observed trends due to climate change. The adjusted hydrology can be used as input to CalSim to support planning studies including the Delivery Capability Report (DCR). This approach of modifying the historical hydrology is preferred by stakeholders compared to generating new scenarios or adjusting higher-order statistics such as drought frequency and duration. The goal is to provide a more realistic baseline for the annual deliveries that contractors can expect under current and near-future conditions (next 5-10 years). This information is typically focused on the long-term average and single dry-year delivery. The DCR also includes an estimate of future climate conditions in 2040, which is outside the scope of this review.

There are many different methods that could be used to perform this adjustment and all have their advantages and disadvantages. It is most important that the changes can be justified in terms of significant physical trends in precipitation and temperature that have already been observed without introducing statistical artifacts. The method selected in this report, runoff curve year-to-month (RC-YTM), adjusts the historical period (1922-2015) based on the recent reference period (1992-2021) for the variables and watersheds that show significant trends using a modified Mann-Kendall test. While the impact of this adjustment on CalSim deliveries is not the focus of this report, the impact is likely modest compared to natural variability and other sources of uncertainty in regulation and demand that could influence near-term deliveries.

The approach is an improvement over the current practice of using the unadjusted historical hydrology to report delivery capability. The quantitative metrics used to evaluate the adjustment methods could be presented more clearly to justify the choice of the RC-YTM approach. Several components of the runoff curve fitting method are unclear, and this regression model also introduces additional uncertainty. The final step of mapping the adjustments to all CalSim rim inflows should be tested to confirm that the adjustments are only applied in basins where significant trends have been detected. In future iterations, there are several opportunities to further improve the method: adjusting temperature directly; evaluating the climate adjustments in the context of natural variability and other uncertainties; and quantifying the impact of sub-monthly hydrologic changes on the monthly rim inflows, for example due to more extreme storms and a greater rain fraction of precipitation that are observed with rising temperature.

Charge Question 1

Is this method an improvement over the use of unadjusted historical data (i.e., an assumption that the historical timeseries is stationary) for representing current conditions? Why or why not?

The DCR estimates the current and near-future water deliveries to contractors. The CalSim model runs used to derive these estimates are based on the historical period 1922-2015. The earlier part of this record does not represent recent trends that have been observed due to climate change, in particular the increasing interannual variability of precipitation and the seasonal shift in runoff due to warming. As a result, the delivery estimates in the DCR should be improved by adjusting the earlier part of the record to reflect the observed impact of climate change compared to the use of unadjusted historical data. The motivation for the adjustment is clear, and the result should more accurately estimate the near-term deliveries that will be available to contractors.

The adjustment is also an improvement in the sense that it is consistent with how other CalSim inputs are handled. Water demands and regulations are updated for each DCR and assumed to apply over the full period. By the same logic, the current climate can be applied to the full period, provided that the statistical changes can be imposed without introducing other artifacts. The selected method (RC-YTM) seems to achieve this goal, though there are some unclear aspects of how the method works and how the performance is measured against alternatives (Charge Question 2). This approach does assume that the historical timeseries is stationary, but it adjusts the statistical properties of that stationary distribution. The detrending of nonstationary hydrology is beyond the scope of the current report but could be a useful future extension.

While the method is an improvement over using the unadjusted historical hydrology in the DCR, it is likely an incremental improvement in terms of communicating the uncertainty and variability in near-term deliveries to contractors. Scenario-based estimates using either GCM projections or synthetic sampling could provide a more complete estimate of uncertainty, though these introduce their own assumptions (Charge Question 5). Regulatory uncertainty could also have a significant influence on these estimates. There are understandable tradeoffs involved, as the historical adjustment requires fewer assumptions and is more straightforward to communicate to stakeholders.

Charge Question 2

How well does the new method account for statistically significant trends to represent a quasi-stationary current climate while avoiding bias or trends that are artifacts?

The proposed method (runoff curve year-to-month, RC-YTM) is a statistical approach to adjust the historical hydrology for variables and watersheds where significant trends are detected. The method contains several steps: (1) mean distance scaling to adjust the interannual standard deviation of precipitation; (2) runoff curves, a quadratic regression model to map annual precipitation to runoff; and (3) a monthly runoff shift based on whether the year is wet, average, or dry. The method is compared to several alternatives and found to perform well according to quantitative and qualitative metrics. There are a few ways that the method and the evaluation framework should be clarified to ensure that no artifacts are introduced.

The methods are judged based on their ability to minimize the relative error in the metrics from the reference period. These metrics are aggregated into a Euclidean distance score with an ideal value of zero. This does not seem to be a true z-score from a standard normal distribution and instead would be better named distance score. The distance score is the main evaluation metric for the adjustment methods, though other qualitative evaluations are also applied. The presentation of the distance score and metrics (e.g., in Table 7) makes it difficult to determine which adjustments are intended and which are artifacts. The charge question would require evaluating both separately, especially because the adjustments are applied differently for each watershed and variable. It would be possible to separate the distance score into two components: the distance for the intended adjustments, and the distance for the metrics not being adjusted. The first would show the ability to account for significant trends, and the second would quantify any artifacts that are introduced.

It could also be useful to present a distance metric aggregated over all annual, monthly, and seasonal metrics to compare the methods directly. There are many metrics, including other PDF/CDF metrics described but not shown in the report, and it is not possible or desirable to present all of them. However, the main quantitative measures that were used to select the RC-YTM method over other approaches should be presented more clearly.

The significant trends are identified using a modified Mann-Kendall test, which is an appropriate choice for timeseries with autocorrelation. The results of the trend tests, namely an increased interannual variance and a seasonal runoff shift, are well supported by literature. However, there is some inconsistency between this test for a continuous trend over time and the way that the adjustments are applied between two discrete periods. It would be a useful future extension to make these consistent either by (1) testing for significant differences using a two-sample test on the discrete periods, or (2) applying continuous adjustments over time to account for nonstationarity. It would also be useful to identify the impact of the choice of a 30-year window for the trend tests, which aligns with the NOAA 30-year climate normal but could lead to different results than a 20- or 50-year window.

One aspect of the method that could introduce bias is the runoff curve. This is a statistical hydrologic model subject to many of the same concerns that the report describes about other hydrologic models used in previous studies (SWAT, VIC, SAC-SMA). The regression from annual precipitation to streamflow introduces additional uncertainty. In Appendix Figure B-1, it seems that the residuals of the regression are on the order of the adjustments made to the annual runoff values. The potential influence of this uncertainty should be investigated, because some adjustments may not be significant compared to the distribution of residuals. This may partly offset the advantage of the method, adjusting precipitation along with runoff in a physically consistent way.

There are several unclear points about how the runoff curve is fitted.

- A different regression is performed for each year of the reference period by resampling nearest neighbors to develop the data used in the regression. Why is this approach used instead of fitting one curve for all data in the reference period?
- Among this set of regression models, any differences in the fitted parameters should be reported, along with the physical interpretation.
- The regression R^2 should be reported for each watershed.
- The regression is fitted to the 30-year reference period 1992-2021. However, this does not allow the case where there is no significant change in runoff efficiency between the historical and current period, where the regression would be fitted to the 1992-1992 period. This choice should depend on the outcome of a significance test to be consistent with how the other adjustments are handled.
- The nearest-neighbor sampling selects 25 values from the 30-year reference period. Are these sampled with replacement? If so, each regression is likely

based on only a few data points. If not, then almost the full period is used for each regression.

- The regression is performed on the log-transformed precipitation. Should the runoff also be log-transformed to avoid negative values? It is possible the y-intercept is constrained to be positive, but this is not stated.

The monthly shift adjustment step is applied differently depending on the year type (wet, average, or dry). The monthly shift is shown to be a significant trend in the first section of the report. However, the separation by year type is not discussed outside of this RC-YTM method. Do the significant trends in the monthly runoff shift also occur in all year types, and what is the physical reason to expect the shift to occur differently depending on the year type?

While the method seems to perform its intended goal, the two aspects of fitting multiple runoff curves and dividing the monthly shifts by year type are perhaps overly complex compared to what the relatively short record can support. It is possible that the method would perform similarly with only a single regression, and a single monthly shift. If this was investigated under another method and ruled out in favor of the RC-YTM approach, the point should be clarified.

Another possible source of bias is how the adjustments are applied beyond the five key basins for which trends are tested (Shasta, Oroville, Folsom, Don Pedro, Millerton). The trends must be mapped to all rim inflows for CalSim3. This is done by combining the rim inflows into 24 groups. However, this could mean that adjustments are sometimes applied to rim inflows that did not show a significant trend. Also, not all rim inflows are included in the calculation of the distance metric, only the five key reservoirs. It is possible that artifacts are introduced in the full set of rim inflows that are not visible using the current evaluation framework.

The method is applied to adjust all variables with significant trends. However, it is not clear that RC-YTM is designed to perform all of these adjustments. For example, two of the five watersheds (Tuolumne and San Joaquin) showed significant trends in the annual mean runoff. The mean distance scaling step in the RC-YTM method is designed to adjust the standard deviation, but not the mean. Does a trend need to occur across all watersheds to be considered for the adjustment?

From the timeseries shown in Figures 17-22, the adjustments seem reasonable and consistent with the trends identified in the first section of the report. In two places there are results that could be minor artifacts of the adjustment. First, in Table 7 the annual FNF change for the RC-YTM method shows a negative standard deviation adjustment, which does not align with the significant increasing trends shown in the first section of the report. Second, in Figure 21 during the dry period 1929-1934 some of the monthly high flows are increased by the adjustment. This may be consistent with the method, but it could also have consequences in the context of how the DCR is interpreted if it reduces the overall severity of this drought period. In general, the impact of the adjustment on dry years should be examined more systematically given how the DCR is used by stakeholders. From the description of the method, it is not clear how these two specific results would occur, which may be a point for clarification.

Charge Question 3

What specific investigations or improvements should be considered in future updates of this dataset?

The changes would be better justified with an explicit link to adjusted temperature. The temperature is the most significant trend across all watersheds (Appendix A), which leads to the monthly runoff shift. This also explains why the monthly shift is more pronounced for the Sacramento tributaries compared to the higher-elevation San Joaquin tributaries. The temperature adjustment would yield several advantages:

- The temperature is a required CalSim input. Adjusting temperature directly, along with precipitation from the current method, would allow physically consistent climate changes to be mapped to runoff changes with less potential for statistical artifacts.
- It would allow simplifying the methodology of the RC-YTM method, for example where the water year type is used to capture different temperature-dependent runoff mechanisms. Instead, these would be included in a hydrologic model, either statistical or physical. The hydrologic model introduces its own error, but the question is unavoidable if the adjustments for precipitation, temperature, and runoff all must be physically consistent.

It may be possible to apply the adjustments using a rolling approach to remove nonstationarity during the historical period. The runoff data in 1922 is less representative of current conditions than 1992. In the current approach, the aggregate statistics of the historical period are adjusted, but the nonstationary trends remain. This question also relates to the choice of trend tests. The modified Mann-Kendall test could be used to show that the adjustment method has detrended the historical runoff.

The changes in CalSim deliveries due to these runoff adjustments could be measured against other potential influences on deliveries, such as natural variability, environmental regulations, and uncertainty from the hydrologic model (runoff curve), as well as uncertainty in the CalSim model itself. This would provide more context for stakeholders to interpret the impact of the runoff adjustments on the DCR compared to other factors influencing deliveries.

The CalSim rim inflows are on a monthly timestep. However, some of the projected impacts of climate change on runoff will occur at the sub-monthly scale. Precipitation events are expected to become more extreme due to Clausius-Clapeyron scaling, and daily runoff peaks will become more extreme due to the increased rain fraction of precipitation (Siirila-Woodburn et al., 2021). This could change deliveries to SWP contractors in two ways: decreased Table A because reservoirs must maintain winter flood pool requirements, but increased Article 21 deliveries as large flood events are released downstream. The monthly rim inflows could be further adjusted to incorporate these changes by estimating the fraction of storable inflow from the daily timeseries and looking for trends in these values over the observed record.

To analyze this difference, Figure 1 shows an example using CMIP5 runoff projections from USBR (Brekke et al., 2014), which could be updated using more recent CMIP6 projections. The fraction of reservoir inflow volume contributed by flows greater than the 90^{th} percentile is compared between the future (2050-2100) and historical (1950-2000) periods for an ensemble of climate models. The 90^{th} percentiles are computed for each model and period to reduce the effect of model biases. For the Sacramento basin reservoirs, a median of 10-15% more of the total inflow volume is expected to come from flows above the 90^{th} percentile. This effect is stronger for RCP 8.5 than RCP 4.5, indicating the influence of increasing temperatures. The increase is less evident for the San Joaquin reservoirs, likely due to higher elevations reducing the impact of rising temperatures.

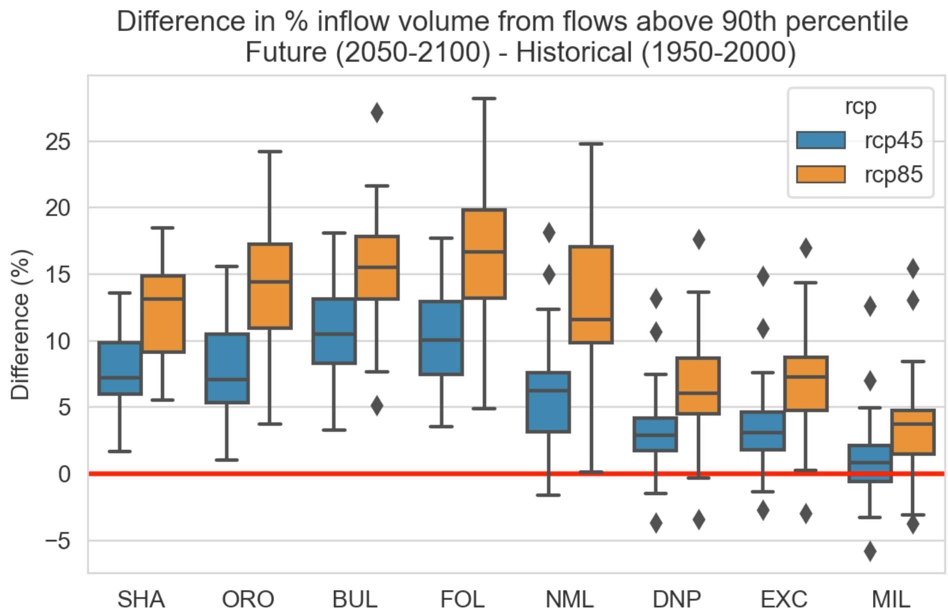

Figure 1: Reservoir inflow volume contributed by daily flows above the 90th percentile, difference between Future (2050-2100) and Historical (1950-2000). Ensemble CMIP5 projections from USBR (Brekke et al., 2014). Eight major reservoirs on the Sacramento-San Joaquin listed from north to south using the CDEC site codes.

The point is that climate change will modify the distribution of daily inflows, and this will have consequences for the volumes that can be stored in reservoirs and delivered on monthly and seasonal timescales. Stakeholders using the DCR to estimate the availability of Article 21 deliveries will need to understand changes to the frequency and magnitude of flood events at the daily scale that cannot be fully analyzed with the monthly record of rim inflows. This also relates to recent research on the availability of high-magnitude flood flows for groundwater recharge (Kocis and Dahlke 2017; DWR 2018).

There may be a benefit to aligning this adjustment method with the future scenario(s) included in the DCR. The 2040 scenarios could extend the same trends from the reference period used in the adjustment. It will take some effort to communicate the scenarios and their assumptions clearly in the DCR – historical, historical adjusted for observed climate change, and projected future climate change. This is not to say that the 2040 projections should be used in the historical adjustment (it is probably better to keep them separate), but that the 2040 scenarios could be developed consistent with the historical observed trends.

Charge Question 4

How frequently should DWR consider updating this dataset?

Other CalSim inputs (demand, regulations, infrastructure) are updated for each DCR every two years. The runoff adjustment method is statistical and only requires observed data, so it should be efficient to implement. There does not seem to be a reason why the runoff adjustment could not also be updated for each DCR. This would be consistent with the other inputs in which the current conditions are applied to the full historical record.

One potential issue with frequent updates could be the addition of new outlier observations that change the significance of the trend tests. It would be interesting to see if the same trends hold with WY 2023 included. The unadjusted historical deliveries should be included as a baseline in every update to compare with the adjusted version, with particular attention to dry year deliveries since the outlier years are the most susceptible to artifacts as the observed data changes.

If the same approach is applied for future updates, at some point it will be adjusting historical data that has already been influenced by climate change. This may already be the case in the current approach in the latter half of the historical period (1960s-1990s). This issue could be addressed with a detrending method to remove nonstationarity rather than adjusting two separate periods.

Charge Question 5

The draft Climate Adjusted Historical Hydrology dataset presented for review is adjusted to a 1992-2021 climate condition. This period is entirely retrospective. With a goal of more accurately simulating the range of hydrologic variability under current climate conditions, what are the pros and cons of taking a more prospective approach in future iterations by, for example, including modeling of potential future conditions to capture a 30-year climate period centered on the current year rather than concluding with the current year?

The goal of more fully capturing the range of hydrologic variability is distinct from the question of more accurately simulating the near-term future (15 years) precipitation. Simulating the range of variability would provide useful information to stakeholders, even apart from the impact of climate change. This could be achieved in a stationary record by sampling synthetic scenarios to represent the range of natural variability in deliveries. It would be interesting to know whether 30-year synthetic scenarios sampled from the same period would show ranges in key statistics on the order of the differences between the reference and historical periods used in this report.

For example, Figure 2 compares the coefficient of variation of annual streamflow for 100 30-year periods sampled from a Thomas-Fiering (AR1) model. The synthetic samples use the full period 1922-2022 and are compared to the two observed periods 1922-1952 and 1992-2022. The full natural flow data comes from CDEC for the eight gages in the Sacramento-San Joaquin. The synthetic sampling assumes a lognormal distribution and does not account for spatial correlation between basins, so this analysis could be improved with a more complex stochastic model. Figure 2 confirms the increase in the standard deviation between the two observed periods, which is also found in the report and is consistent with increasing interannual variability under climate change. However, a similar range occurs with 30-year synthetic samples based on the full record, which underscores the importance of natural variability in estimating these statistics. This may be less of an issue at the monthly scale, as the monthly runoff shift is a direct result of rising temperatures and would be unlikely to occur in stationary synthetic simulations.

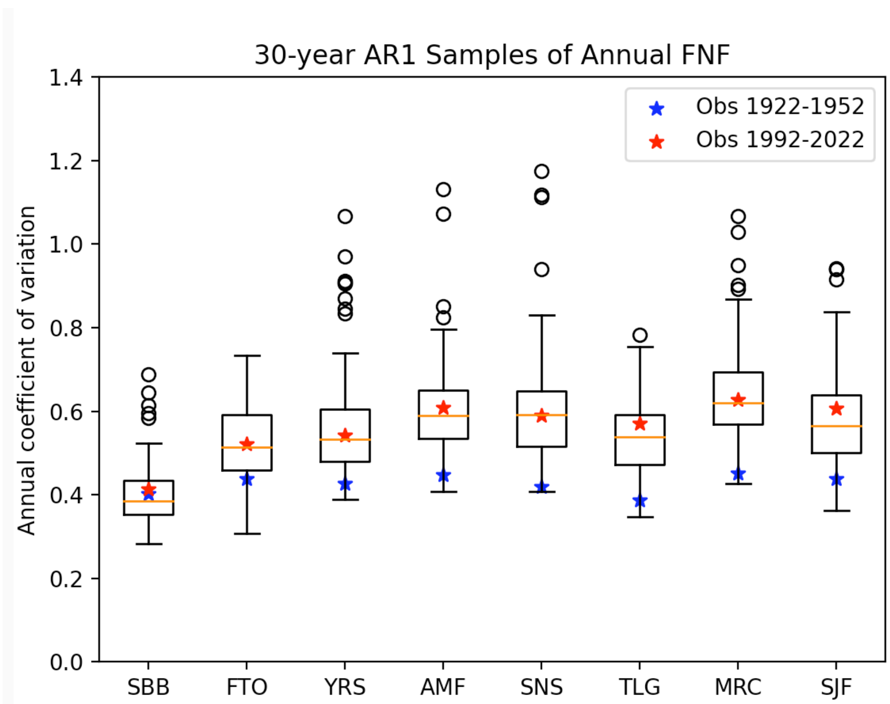

Figure 2: Coefficient of variation of annual streamflow for the eight major basins on the Sacramento-San Joaquin, named by the CDEC FNF site codes. 100 synthetic samples of 30-

year periods from a Thomas-Fiering (AR1) model compared to two observed 30-year periods.

Modeling potential future conditions is a useful research goal that could support the DCR. One potential challenge is that the modeling methods would introduce additional assumptions and uncertainties. These would need to be clearly communicated to stakeholders to justify replacing or augmenting the observed data. The modeling would come from either GCMs or synthetic scenarios of precipitation. In the next 15 years, both would be dominated by natural variability. The GCM projections would also contain substantial uncertainty in the choice of model and downscaling method (Lehner et al. 2020, Lafferty et al. 2023). In either case, the range of the projected precipitation would have little to do with the degree of climate change on this near-term scale. GCM projections would likely also require a bias correction to line up with recent observations, leading to circular logic of adjusting the future projections based on the observations only to then perform the reverse operation. A synthetic scenario approach could mitigate this problem to some extent. However, it would require extrapolating observed trends beyond the record. This is a reasonable assumption but a difficult one to support quantitatively.

The question comes down to whether precipitation projections for the next 15 years would better reflect current climate conditions than the 15 observed years that would be removed from the window (1992-2007). Given the uncertainties involved, this may not be the case. However, if it were framed as sampling synthetic scenarios, rather than a projection, the experiment would at least provide a more complete view of natural variability. The synthetic scenarios could be modified bottom-up based on the observed statistics that show significant trends in this report without extrapolating the trends into the future.

A more prospective approach may also need to consider the changes in operations and regulations that could result from climate change or other factors. While the CalSim model runs would provide a baseline estimate of deliveries assuming current operations, in the current setup it would not allow the system to adapt to changing hydrology. This effect may be negligible if the hydrologic scenarios do not deviate too much from the historical range. It could become more significant if a wide range of scenarios are tested.

In the current DCR model runs, there seems to be a communication benefit to separating the historical scenario from the modeled future scenario. Stakeholders primarily rely on the historical scenario, and they can also interpret the 2040 scenarios with appropriate caveats. However, the proposed approach would integrate the future scenario information into the historical adjustment, rather than vice versa. This could create more of a challenge to communicate the methods and uncertainties involved in the historical adjustment.

References

Brekke, L., Wood, A., and Pruitt, T. (2014). Downscaled CMIP3 and CMIP5 hydrology projections: Release of hydrology projections, comparison with preceding information, and summary of user needs. National Center for Atmospheric Research and U.S. Bureau of Reclamation.

CA Department of Water Resources (2018). Water Available for Replenishment: Final Report.

Kocis, T.N., and Dahlke, H.E. (2017). Availability of high-magnitude streamflow for groundwater banking in the Central Valley, California. *Environmental Research Letters*, *12*(8), 084009. DOI: 10.1088/1748-9326/aa7b1b

Lafferty, D.C. and Sriver, R.L. Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6. *Authorea Preprint.* April 30, 2023. DOI: 10.22541/essoar.168286894.44910061/v1

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491-508. DOI: 10.5194/esd-11-491-2020

Siirila-Woodburn, E.R., Rhoades, A.M., Hatchett, B.J., Huning, L.S., Szinai, J., Tague, C., Nico, P.S., Feldman, D.R., Jones, A.D., Collins, W.D., and Kaatz, L. (2021). A low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth & Environment*, *2*(11), 800-819. DOI: 10.1038/s43017-021-00219-y