SEPT 2023

Independent Peer Review of the State Water Project – Delivery Capability Report, Part 2

An individual letter review for the Delta Science Program

Prepared by

Dr. Jon Herman – University of California, Davis

Delta Science Program

DELTA STEWARDSHIP COUNCIL

Table of Contents

Summary

This report describes a method to develop risk-informed future climate scenarios to support the SWP Delivery Capability Report (DCR). While previous DCRs have relied on a single climate scenario, the goal of this work is to better capture the range of uncertainty in SWP deliveries 20 years in the future due to climate change while maintaining the transparency and utility of the DCR for local planners. The approach combines top-down scenarios from global climate models (GCM) with bottom-up sampling of climate perturbations to create a response surface of system performance. A probability distribution fitted to the GCM scenarios is then applied to the response surface to estimate a distribution of system performance metrics and identify thresholds of concern, such as the 95% non-exceedance value.

The proposed approach provides substantially more information about the range of potential future conditions than the previous approach, which provides mean and dry-year delivery estimates from the historical record and a single future climate scenario. This method has the advantage that it can inform stakeholder decisions based on different levels of risk aversion or alternate conceptions of climate uncertainty. The steps taken to estimate the probability of different levels of climate change and sea level rise are reasonable within the limitations of existing GCMs. The report is careful to note that the true probabilities of these changes are unknown, but that in order to estimate risk for stakeholders it is necessary to treat the GCM projections as independent samples from a distribution in the absence of other information. There are a few points where the use of climate change scenarios can be clarified, but in general the approach uses the best available information and combines the respective advantages of top-down and bottom-up methods to estimate climate risk.

One point where the approach can be improved in future DCRs is the treatment of natural variability. Estimating the distributions of system performance metrics in the future depends on both the probability of climate change as well as natural variability. While the approach is thorough on the first point, the natural variability may be underestimated for several reasons, including averaging across climate realizations, averaging performance metrics across time, and applying perturbations to the historical record rather than using climate scenarios directly or generating synthetic sequences of weather and hydrology. All of these steps are justified in the context of the current DCR, and in some cases are specifically requested by stakeholders (e.g. the objective that the scenarios should follow the historical pattern of wet and dry years). However, they require that the historical

record contains an adequate sample of natural variability. These are areas that could be improved in the future to provide a more complete estimate of system risk. There is growing interest in this type of large ensemble approach, and the Working Group is well-positioned to take on this challenge.

Charge Question 1

Is the procedure developed by DWR appropriately documented? Is there anything missing from the documentation?

The procedure is mostly clear. Two main points could be improved. First, while the stress test example using SAC-SMA and CalLite is well documented (Section 5.3), it could be more clear how this process will be integrated into the DCR using the planned CalSim 3 results. Aside from the VIC model described in Section 5.6, other aspects of the experiment may be different. As one example, the stress test results focus on the 8-River Index as an important system metric. However, the index does not require running CalLite or CalSim, so the metric(s) of interest for the DCR would likely be different. This may be a question of structure, as the results of the CalLite study are presented before the methods needed for CalSim. The section order may benefit from having a full description of the methods first, followed by example results from the CalLite study to show the benefits of the approach.

Along these lines, it would also be helpful if the methods flowchart in Figure 5-2 were revised to follow the order of the sections in the report. For example, the LOCA2 scenarios and distribution fitting are not listed in the flowchart, though they are implied in Step 2. Step 2 also indicates that the level-of-concern (i.e. system performance) is needed, but if I understand correctly, this requires running the system model first (Steps 4-6) unless a hydrologic metric such as the 8RI is chosen rather than a system performance metric. There could also be separate methods flowcharts in the report for the current and future DCRs to show how the approach may be further improved.

The second main point to clarify is the use of the stochastic weather generator (WGEN) in this approach (Figure 5-2, Section 5.5, and throughout). As the report describes, the WGEN is a powerful tool that can create daily sequences of precipitation and temperature based on clustered weather regimes. However, the method is used in a limited way in this study. Three factors are used to perturb the historical hydrology: annual temperature change, annual precipitation change, and Clausius-Clapeyron scaling. These are statistical modifications of the observed record with a physically-based justification, but they do not involve stochastic generation. Also, because the hydrology is aggregated to a monthly timestep, the Clausius-Clapeyron scaling may not be very influential. This is already documented in the report, and the use of the full WGEN is a focus for future work (Pg. 5-43). However, it should be clear throughout the report that the current DCR is not using synthetic scenarios. It would be useful to know what are the barriers to getting there, because the weather generator would provide a much more complete sampling of natural variability while matching the statistical properties of the observed precipitation (Najibi and Steinschneider, 2023). Page 5-3 indicates that stakeholders are gaining interest in this type of large ensemble approach.

(Section 5.2) The approach of selecting LOCA2 scenarios, watershed averaging, and flow-weighted averaging to obtain a domain-wide average signal are clearly described. However, it is not clear why a domain-wide average is needed, other than the response surface. The spatial averaging is already mentioned as a limitation (Page 5-40 and Chapter 6) as it removed spatial variability. It would be useful to test the impact of this assumption, for example by checking the correlations between the projected climate change in each basin and the domain-wide average. There does not seem to be a technical reason why the LOCA2 and WGEN scenarios could not be used to perturb the spatially distributed historical climate before aggregating, rather than after.

(Section 5.3) It would be useful to know what drives the choice of SAC-SMA or VIC for these studies (accuracy, data availability, runtime), and to what extent this choice affects the selection of scenarios. The same question applies to the choice of system model, CalLite, CalSim, or WEAP. Appendix B shows that different combinations of these models have been used for recent climate change studies in California.

(Section 5.2.5 and 5.3) More information about the choice of sample size would be helpful. This applies to both the sample from the bivariate distribution (10,000 samples) and the sample size used to generate the response surfaces (Figure 5-12). This large sample size for the bivariate distribution is probably only needed if the weather generator is being used to develop new sequences of precipitation and temperature. If these samples are only used to perturb the historical record, a coarser sample is likely adequate because it is not sampling natural variability. Also, the resolution of the sample grid for the response surface may need to align with the choice of the sample size of the bivariate distribution, otherwise the response surface will be interpolated over a grid that may be too coarse.

(Section 5.6) It is not clear if the WGEN output in this section refers to the perturbed historical precipitation and temperature or synthetically generated traces planned for the future. In either case, Step 4-5 (perturbing the historical streamflow using the VIC output) may not be needed, because the potential bias in GCMs will have been avoided.

(Section 5.6) The hydrology perturbation method involves monthly ratios of the future to historical GCM scenarios. Clarify whether this is the same process in the CalSim 3 report (Page 20-6), which involves monthly quantile mapping to better capture the extremes.

Charge Question 2

Does the procedure apply rational and defensible evidence for the steps taken and techniques used to capture the probability of projected changes related to climate and sea level rise? Why or why not?

The report is careful to note that the GCM scenarios are not able to be placed in a probabilistic context (Page 5-2), because the true probabilities are unknowable. However, the method does require treating the GCMs as independent samples from a bivariate distribution. The goal is risk analysis, so the concept of probability is unavoidable. The focus on the term level-of-concern highlights this caveat. The method is not tied to a single estimate of climate uncertainty, but instead allows modifying the probability distribution with alternate conceptions of climate uncertainty (Section 5.4). It is possible that these estimates or scenarios could include other statistics of interest beyond the mean annual temperature and precipitation.

The method is designed to estimate the probability distribution of system performance metrics under climate change. This requires identifying both the probability distribution of system performance in the absence of climate change, as well as the probability of different impacts of climate change. The first point relates to natural variability and a full understanding of system risk (Charge Question 3). The second point applies to this Charge Question. These two goals are sometimes at odds: identifying the probability of climate changes may require removing natural variability, but the probability of future water supply outcomes would need to include both climate change and natural variability.

The method captures the bivariate probability distribution of annual changes in mean precipitation and temperature, assuming that each GCM scenario is an independent sample from a Gaussian distribution. This is a reasonable assumption in the absence of other information. However, because the approach focuses on the annual average change, it may not directly capture the GCM-based probabilities of other climate changes that are important to water management such as the frequency and severity of droughts. Whether these differences are due to climate change or natural variability would be difficult to separate. The differences may be even more strongly influenced by other factors including the choice of GCM and downscaling method (Lafferty and Sriver, 2023). The impacts of climate change must be aggregated at some level, because as the report notes, providing thousands of scenarios to explore each dimension would be untenable for stakeholders. It could be tested whether perturbing the historical record with the mean precipitation and temperature will approximate the probabilities of changes in extremes that occur in the GCM scenarios with similar average properties.

One minor unclear point is whether the Clausius-Clapeyron scaling factor should be included in the distribution fitting. It is part of the response surface experiment but does not seem to be assigned a probability from the GCMs. However, the influence is likely small for the monthly timeseries.

Sea level rise is treated separately from regional climate using a defensible approach. The SLR estimates are restricted to near-term (before 2050) when there is higher confidence in the projections. The analysis assumes that the SLR scenarios and the climate scenarios are independent. If they were correlated (e.g., if the warmer local scenarios also lead to higher SLR) this may affect the estimates of the joint probability. Section 5.7 (Pg. 5-50) suggests that SLR within this projected range will likely have only a modest impact on near-term Delta operations relative to climate variability and change.

Charge Question 3

Do the new scenarios provide enhanced information for water users about potential future conditions and system reliability risks? If not, why?

The proposed approach provides substantially more information about the range of potential future conditions than the existing DCR approach, which provides mean and dry-year delivery estimates from the historical record and a single future climate scenario. The new scenarios can inform stakeholder decisions based on different levels of risk aversion or alternate conceptions of climate uncertainty.

The approach is focused on system reliability risks due to climate change, which are reasonably captured within the limitations of GCM-based estimates of precipitation (Charge Question 2). However, there are several points in the methodology that remove or omit natural variability, which could result in underestimation of the full distribution of system risks:

- The LOCA2 scenarios with only one realization are removed (Table 5-2) because of the difficulty in separating climate trend from variability. This is reasonable for the goal of estimating the probability of climate-driven changes by removing stochasticity (e.g. Page 5-20), but the most complete picture of system risk would include both the climate trend and the variability.

- The climate changes are also separated from natural variability by taking a linear trend on the 30-year average precipitation. This may reduce the variance in the annual precipitation values in the perturbed historical scenarios relative to the GCM scenarios. This would not be an issue if the GCM variance is attributed to bias, but it could be tested.

- The remaining GCMs are averaged across realizations before fitting the bivariate normal distributions (Page 5-15 Step 4). The result will represent the system risk level not exceeded by (e.g.) 95% of the sampled scenarios, but if the scenarios represent the average of 1-3 realizations, is it possible that this variability could change the estimate of the non-exceedance probability. Figure 5-7 is clear on this point. It shows that while the variance in temperature depends on the SSP, the variance in precipitation instead depends mainly on the choice of GCM and the realization. As a result, the distribution fitted without averaging the GCM realizations has a much larger variance in precipitation (not in temperature). Because the variance is not attributable to climate change, it is removed from the analysis.

- The variability in annual performance metrics is removed by averaging over time. For example, the CDF in Figure 5-15 shows the distribution of the expected 8RI over the scenarios, but it should be clarified that this is not an annual exceedance curve. In the context of the DCR, it could be useful to include both the distribution of expected changes averaged over time, as well as the distributions of the performance metrics over all years in the simulations. This would be consistent with the historical dry-year analysis currently included in the DCR, as well as other recent bottom-up climate studies of California water supply (e.g., Ray et al. 2020). Part of the difference might be that the scenario samples drawn from the probability distribution are used to find the expected value on the response surface, but if I understand correctly, they are not run through the system model to obtain the timeseries of performance metrics.

The last points of the approach that may underestimate natural variability are the two perturbation steps. The sampled climate scenarios are first used to perturb the historical precipitation and temperature, and then the hydrologic model output is used to perturb the historical hydrology. The first part of the approach (perturbing precipitation and temperature) was developed for this study. It is justified in terms of stakeholder interpretability, and it is clear from Section 5.5 that expanding the set of stochastic precipitation and temperature scenarios is a priority for future DCRs. The second part of the approach (perturbing hydrology) comes from past CalSim studies (Appendix B, and Charge Question 4) so it is a bit outside the scope of the review, but it has been a critical step in many other climate change studies.

Analyzing only the historical sequence of drought events with perturbation methods will likely omit some of the natural variability within and across GCM scenarios. This would also almost certainly be the case for scenarios created by the weather generator if they are used in future DCRs. For example, Najibi & Steinschneider (2023) show a wider range of extreme drought events from the WGEN, even in a stationary scenario without climate perturbations. Figure 11 shows results for the Tuolumne watershed, where the minimum precipitation over 1-10 year periods decreases by 15-25% when sampling a stationary 1000-year record compared to the 100-year observed record. This is larger than the decrease in the GCM realizations in this report (Figure 5-7), as well as the range of scenarios tested for the response surface (Figure 5-12). This approach comes with the caveat of extrapolating extreme events from a shorter observed record to a longer synthetic record, but it aims to do so in a physically-based way. This point underscores the importance of natural variability in the risk to deliveries.

For this report, it would be helpful to add a discussion of the reasons for maintaining the perturbation steps in the current approach, as well as what improvements in the scenario modeling chain would need to be achieved before their ensemble output could be used directly in CalSim 3. It could also be useful to see how the stress test results, for example the response surface (Figure 5-12) and CDF (Figure 5-15), would change if a different climate scenario were used in place of the historical record, or if all GCM realizations were included in the distribution fitting. In both cases it might be anticipated that the distribution of system performance would have a larger variance, which could affect the calculation of the 95% level-of-concern for the 8-River Index and other metrics.

In future DCRs if the WGEN scenarios are generated to create the response surface, even a large sample of natural variability would likely be averaged out by the interpolation step. There could be two ways around this: either use the WGEN to generate scenario samples directly from the bivariate distribution and run them through the systems model (if computation allows), or create the response surface using both the mean and variance to represent natural variability. This comment would only apply to future work where the WGEN is used to create new sequences of precipitation and temperature.

A last point relates to the choice of system performance metric to determine risk. In the stress test example using SAC-SMA and CalLite, the April-July 8-River Index (8RI) was chosen as the key metric by stakeholders. The report mentions many good reasons for this choice, but I would also expect it to be very sensitive to snowpack decline—for example, if the same amount of annual precipitation falls as rain during November-March, it would not be included in this metric, but would still provide water supply. The interpretation would be different if a lower 8RI occurs because of earlier runoff versus less total runoff. Using the annual 8RI instead would provide the same advantages while avoiding this issue.

The 8RI performance metric only requires running the hydrologic models but does not require running CalSim/CalLite. This would not be the end goal of the DCR, but it could allow wider explorations of the scenario space because it would not require creating all of the CalSim inputs via the perturbation method. This is another advantage of the metric that could be included in the documentation, that it can be used to select level-of-concern scenarios without running the system model for each sampled scenario.

Charge Question 4

Is this procedure an improvement over other previously used approaches to climate scenario selection/development? Why or why not?

Previous DCRs provide delivery estimates for the historical scenario and a single perturbed future climate scenario. The larger set of scenarios developed in this approach is a significant improvement as it provides a more complete representation of uncertainty and leverages the new LOCA2 scenarios. The approach is also an improvement over previous top-down studies of California water supply under climate change (Appendix B) because it provides a full response surface of system performance over a range of potential future changes, recognizing that these are uncertain. The approach is in line with previous bottom-up studies. Finally, this method uses a climate period approach rather than a transient approach, which allows estimated future conditions at any time to be applied uniformly either to the historical record or to a synthetically generated record of a chosen length.

While the scenario selection and generation tools (LOCA2 and WGEN) are state of the art, their application here is limited to perturbing the historical record. As a result, the scenarios used to evaluate the distribution of system risks will likely underestimate natural variability. The climate perturbation step is discussed above. The hydrology perturbation step has been used in many past CalSim studies (Figure 5-2, Step 5), but it is not clear that it is needed in this method.

The reasons for using the perturbation approach include potential biases in the VIC model (Section 5.6). The CalSim 3 report (Page 20-5) mentions biases in the GCMs as the main reason for this approach. It is true that the GCM interannual variability and mean annual precipitation have been shown to be biased in the past (Persad et al., 2020), though this may be improving in the LOCA2 scenarios with updated bias correction methods (Pierce et al., 2023). However, the method developed in this report does not run GCM scenarios through the hydrologic model. Instead, they are used to fit a distribution, from which many samples of synthetic weather can be sampled, either by perturbing the historical climate or generating novel sequences from the WGEN. In both cases the GCM bias issue may be avoided—for example, the WGEN provides a good match to the observed annual mean and standard deviation of precipitation (Figure 8 of Najibi & Steinschneider, 2023). It would be useful for the report to include a discussion about whether this could eliminate at least the step of perturbing the hydrology. If biases remain in the VIC model, rather than perturbing the historical hydrology, it may be preferable to recalibrate or switch to a different hydrologic model such that the output could be used directly in CalSim. An additional possible reason for keeping the perturbation strategy is that direct hydrologic model output could create difficulties for the CalSim solver if constraints that were tailored to the historical record become infeasible. If this is the case, it should be discussed as well.

Charge Question 5

Are there specific investigations or improvements that should be undertaken in future updates of this approach or use of this procedure to develop additional scenarios at time periods further into the future?

The report focuses on a 20-year horizon (2043), which is a good compromise that is both a long enough planning horizon for municipal and agricultural water users, but also sufficiently near-term that model projections have a narrower range of uncertainty. A longer horizon would require navigating a higher degree of uncertainty in sea level rise projections (Figure 5-19) and their influence on Delta operations. This includes not only the uncertainty in global sea level rise, but also uncertainty in how these changes may lead to salinity intrusion in the Delta. It would also introduce much more uncertainty in regulations and infrastructure, and may require estimating how system operations are likely to adapt to different degrees of climate change. On the other hand, a higher degree of uncertainty in the annual temperature and precipitation would be straightforward to incorporate into the existing method, because the system response surface will already have been created. The method would not be developing additional scenarios, but instead conditioning the distribution of system risk on a different timeframe of GCM projections. However, this additional uncertainty in precipitation would likely be driven by the choice of GCM and realization rather than the degree of climate change (as shown in Figure 5-7). Rather than projecting further into the future, more fully capturing natural variability in the present would serve the near-term goal of estimating delivery risks for stakeholders.

Additional Comments

(Page 5-2, 5-3) The comparison of top-down and bottom-up methods states that top-down scenarios cannot be placed in a probabilistic context, but bottom-up approaches can provide probabilistic information. This description does not align with the proposed approach, which uses top-down GCM scenarios to estimate a probability distribution and bottom-up methods to sample a large number of scenarios from that distribution. In other words, the bottom-up methods by themselves do not generate probabilistic information. Estimating a probability distribution from GCMs is imperfect but reasonable in the absence of other information. This is not an issue for the approach as long as appropriate caveats are included—for example, as in Section 5.2.5. I suggest summarizing the same points in the introduction to avoid confusion about the analysis of probabilities.

(Figure 5-2) Clarify that the Weather Generator is only partially used here to scale the historical climate data, not to generate new sequences of weather, as may be done in future DCRs.

(Figure 5-3) Clarify that the method averages the LOCA2 scenarios over the major watersheds, not the HUC8s.

(Page 5-12 Line 4) It is clear how the domain averages are performed, but more explanation of why this step is needed would be helpful.

(Page 5-15, Step 1) Clarify that the 30-year average is taken on the annual precipitation data, not the original monthly data.

(Page 5-22 Lines 7-9): The references to left/right figure panels are switched.

(Page 5-22 Line 20, and Figure 5-8) Clarify that the SSP-driven change is increasing variance between GCMs, not increasing interannual variance in precipitation. The result shows that higher SSPs create more disagreement between GCMs.

(Figure 5-9 and 5-10) The comparison between LOCA2 and IPCC WG1 precipitation change is useful, but the validation of LOCA has been done in more detail in publications and is a bit outside scope here. Because the IPCC data is not quite the same geographic region, the comparison is challenging and could be omitted.

(Figure 5-13) This is a local sensitivity analysis that depends on the choice of the points on the response surface from which the changes are estimated. The results make sense, but this choice should be clarified.

(Page 5-42) Scenarios are perturbed using three factors: average precipitation, average temperature, and Clausius-Clapeyron scaling. But because the scenarios are not a full factorial sample of these three factors, it is difficult to understand which combinations are selected and why. If possible it would be helpful to include the scenarios figure from the WGEN report to clarify this (Figure 4 from Najibi & Steinschneider 2023).

(Table B-3) Clarify whether these previous bottom-up studies used synthetic weather generation or the current approach of perturbing the historical climate.

References

California Department of Water Resources (2022). CalSim 3 Report: A water resources system planning model for the State Water Project & Central Valley Project. Sacramento, CA.

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

Najibi, N. and Steinschneider, S. (2023) A process-based approach to bottom-up climate risk assessments: Developing a statewide, weather-regime based stochastic weather generator for California. Technical Report: Cornell University, Ithaca, NY.

Persad, G.G., Swain, D.L., Kouba, C., and Ortiz-Partida, P. (2020). Inter-model agreement on projected shifts in California hydroclimate characteristics critical to water management. Climatic Change 162, 1493–1513. https://doi.org/10.1007/s10584-020-02882-4

Pierce, D. W., Cayan, D. R., Feldman, D. R., & Risser, M. D. (2023). Future Increases in North American Extreme Precipitation in CMIP6 Downscaled with LOCA. *Journal of Hydrometeorology*, *24*(5), 951-975. DOI: https://doi.org/10.1175/JHM-D-22-0194.1

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M., & Brown, C. (2020). Vulnerability and risk: Climate change and water supply from California's Central Valley water system. *Climatic Change*, *161*, 177-199. DOI: https://doi.org/10.1007/s10584-020-02655-z