CalSim 3 in COEQWAL: Model Purpose, Review, and Limitations

1. Purpose of this document

COEQWAL publishes raw CalSim 3 output alongside summarized results. The raw model files and output are served deliberately, as a transparency and verification feature in support of the summarized results.

The COEQWAL project has applied the CalSim 3 model for a somewhat different purpose than most agency modeling efforts, and readers may therefore evaluate the results against expectations formed by those established applications. This document describes the ways in which the COEQWAL CalSim 3 modeling process may diverge from agency methods. The document also describes the types of discrepancies or errors that may arise in a CalSim 3 modeling endeavor as well as the ways in which the project team tried to minimize them. The purpose of the document is to provide the reader with additional context around the development of CalSim scenarios to guide technical evaluation and use of the data published by the COEQWAL project.

2. The same model, applied differently

CalSim 3 is the planning model developed and maintained by the California Department of Water Resources and the U.S. Bureau of Reclamation for the State Water Project, the Central Valley Project, and the broader Central Valley water management system. It is the model used and referenced most frequently in California water-management allocation planning. COEQWAL uses the same model; what differs is the application.

In agency application, individual model runs support specific regulatory and legal processes in which the absolute values carry consequence (e.g. DWR's delivery capability reports for State Water Project contractors), sometimes in comparison to a small set of proposed alternatives. Runs are developed and refined over extended periods; draft results are shared with parties to the process, feedback is incorporated, and the cycle repeats until the final run can withstand technical scrutiny in that application. This process fits its purpose: a small number of runs, examined deeply, whose absolute values have received some level of buy-in from parties most invested in the narratives the values support.

Participating in that refinement cycle requires the ability to run, audit, or independently evaluate CalSim itself, which takes specialized skills and sustained resources. This means the feedback process reaches the limited set of organizations equipped to engage with the model directly and persuasively. Most water users within the model's geographic domain are not in a position to participate in it.

COEQWAL's purpose is different: a library of over 100 scenarios spanning many management strategies and hydroclimates, built so that multi-sector outcomes can be compared across them, and addressed to an audience with little (or no) experience running planning models like CalSim 3. The goal of making the COEQWAL library accessible is to build a broader understanding of how California's water management system works and to help populate the landscape of possible options for the future. Put more directly, the purpose of COEQWAL scenarios is to help illustrate the patterns, trends, and qualitative tradeoffs that different interventions may produce under a range of hydroclimate conditions, not to establish policy recommendations or operating criteria for agencies.

For a library of that size and an audience of that breadth and a timeline as abbreviated as COEQWAL's, the agency process used to develop and publish CalSim 3 scenarios does not readily translate: extended detailed scrutiny and iterative refinement does not scale across the ensemble, and a time-intensive expert feedback loop cannot be the mechanism by which the intended audience engages. Instead, the scenarios are developed following a more conceptual approach informed by input received from the various committees and participants through the outreach and discussion process facilitated by the project. This approach – namely that not every aspect of each scenario needs precise tuning and specification to provide useful information via model results – allowed ideas to be propagated into model runs within the timeline of the project. All model results were thoroughly reviewed, checked, and refined as needed before being made available on the COEQWAL project website. However, this review process was tailored to the project timeline and purpose the model scenarios are meant to serve.

Recognizing that this is context for model development may be insufficient for some traditional consumers of CalSim models, the COEQWAL project team has provided tools to explore data summaries, documentation to understand and reproduce the derived results, and access to the raw model files and outputs. The model files and results are provided by COEQWAL in the spirit of openness, transparency, and trust that users who engage in detailed technical review will do so in a good faith effort to improve the models in support of COEQWAL project goals. However, because COEQWAL modeling has not undergone the same level of review to which most agency models have been subjected prior to publication, the results and models are not intended to be viewed with the same weight as official agency models. Furthermore, agency involvement in the project (via Steering Committee participation or other activities) does not constitute official approval of the scenarios or results: the actors who are normally assumed to have provided (implicit or explicit) authorization of model files or results (agency modelers and their hired consultants) have not officially approved or otherwise endorsed the COEQWAL scenarios.

The COEQWAL scenario development effort is not intended to replace agency modeling. Runs refined for regulatory application answer questions that COEQWAL's runs are not built to answer. The disclaimer that confronts COEQWAL website users intends to make and reinforce this point. The two applications serve different purposes even though they use the same model.

3. How the published material is structured

The COEQWAL project publishes multiple layers of CalSim-derived data and interpretation:

The raw-data layer. The website serves CalSim 3 output and complete model input files for every scenario on its Data page (coeqwal.org/data). These data and files should be sufficient for a detailed evaluation of scenario implementation, if desired.

The interpretive layer. The Explore section of the website and the summary briefs provide a synthesis of model results using outcome levels aggregated across time and compared across the scenario ensemble under matched hydroclimates. The unit of analysis is the pattern across scenarios, not the value at a single node in a single month. Smaller scale model discrepancies may not be visible at this aggregated resolution.

The derivation between them. The key-outcome methods briefs document how raw values are translated to outcome levels. Raw data, method, and summary are all available for inspection, so any summarized result can be traced back and checked by a reader with the inclination to do so.

The website itself is organized around this distinction. The Explore section, where most readers will spend their time, works entirely in outcome levels: results summarized into a form that does not require modeling expertise to read. The raw model output and the full set of briefs sit on the Data page for anyone who wants to look deeper or to check the summarized results against the underlying values. No reader is asked to interpret raw model output to use the site; the results presented for general use are the pattern-level comparisons. The modeling and review process prioritized resolving issues and errors such that these pattern-level comparisons are valid.

4. Established limitations of CalSim 3

The limitations below are derived from the model's documentation and in the published review literature.

- **Monthly timestep.** The model simulates monthly operations. Sub-monthly variability, including short-duration events, is not directly simulated; monthly values are averages or end-of-month storage conditions by construction.

- **Historical sequence:** CalSim 3 is normally run with hydroclimate inputs (i.e. river inflows to reservoirs, irrigation demands, valley floor meteorology) that follow a single historical sequence. Future climate conditions are commonly applied to this sequence via monthly and annual factor adjustments, but retain the same annual sequence of wet and dry years. This limits the range of year-to-year variability or sequences that are represented in a CalSim scenario. Using the historical sequence directly can also introduce non-stationarity, as the early part of the observational record (starting in the 1920s) is not climatically identical to today's warmer climate. The 2023 DWR adjusted historical hydroclimate (used as the basis for COEQWAL scenarios) corrects this issue.

- **Priority-based optimization.** Operations are represented as stepwise constrained mixed integer linear programming optimization applied to a storage and conveyance network. Some behaviors in the output -- and the reasons for their occurrence -- follow from that structure rather than from any well-defined physical process or specific logical sequence of decisions. This model structure often requires abstraction or reformulation of real-world operating rules, a process that is subject to modeler judgement. Evaluation and interpretation of model results should be made in the context of these characteristics of the model.

- **Fixed operating rules under changed conditions.** Operating rules encoded in the model reflect an interpretation of existing regulations and agreements or hypothetical modifications thereof. The ruleset does not change or adapt through the course of a simulation (e.g. to reflect the timing of past interventions or as a dynamic response to some new conditions). Similarly, the rules are only changed for simulations using future hydroclimates if modelers specifically make such modifications. In this way, current or proposed operations are tested against possible future hydroclimates to evaluate how well different strategies may perform.

- **Level of development and input vintage.** Land use, and the irrigation demands derived from it, is set according to a single static map that is applied uniformly to each year of the simulation. Similarly, facilities and infrastructure components are assumed to be constant across the entire simulation. The degree and specificity to which land use or infrastructure from a particular period is included in model inputs is a subject of model purpose and modeler judgement. Local, small, or recent changes are likely not reflected in commonly used input datasets.

- **Heterogeneous fidelity across demand units.** CalSim 3 represents some demand units more accurately than others, reflecting differences in underlying data availability, calibration attention, and how demands are specified. The project has decided not to quantify or label this per location of interest (LOI), because a single published index would carry false precision of its own; this document is where the issue is addressed in general.

The last item is where a reader is most likely to encounter a mismatch between the model and a place they know. An important distinction to keep in mind is between relative and absolute accuracy. A persistent bias in how a demand unit is represented is largely shared across every scenario in the ensemble, because all scenarios run through the same model structure. Cross-scenario differences, which is what the outcome-level comparisons report, "cancel out" much of that shared bias in a way that absolute values at a single location do not.

5. Types of Discrepancies in CalSim Scenarios

Discrepancies from expectations in CalSim 3 model results – that is, numerical values for specific outcomes at specific locations that are different from what an informed but non-modeler observer might expect -- can be grouped into three broad categories or types. The identification of a discrepancy by a modeler relies entirely on the expectations that that modeler assigns to the model and its results. As these expectations may vary considerably, the proposed typology and implications for discrepancies are likely neither comprehensive nor absolute. For the COEQWAL project, however, the following discrepancy types provide a useful framework to describe how issues have been identified and if/how they were reconciled.

The first type of discrepancy that may appear in CalSim 3 results is a divergence from measured or observed history, often because of the structure and assumptions of the CalSim 3 model. An obvious example is that the application of static land use, infrastructure, and operating rules across a 100-year historical time series will not align with observations of flows (or deliveries, or reservoir storage) over that entire period. Most reservoirs in the Central Valley were not built in the 1920s, so Central Valley river flows will be much different in the model than in historical records! These sorts of discrepancies are an inherent function of the CalSim 3 model and overall planning modeling approach and were not addressed by the COEQWAL modeling team. Additionally, as noted in the previous section, the presence of systematic discrepancies across all scenarios does not preclude evaluation of differences among scenarios from a comparative perspective.

A second type of discrepancy is one that occurs due to a mismatch or error in input datasets or component sub-models. Examples of this type of discrepancy might be a numerical error in a reservoir inflow time series, bias or uncertainty in ANN-estimated salinity under certain flow conditions, or mis-calibrated irrigation demands for a specific irrigation district represented in CalSimHydro. Because they are embedded in commonly used input and model structures, these sorts of discrepancies tend to occur across many scenarios and will tend to introduce a consistent error that may not always be obviously detectable. The COEQWAL modeling team reviewed common model inputs and results (including inflows, demands, Delta salinity, deliveries, and select river flows) for anomalous values that may have indicated a data or sub-model issue. This level of review would not identify less obvious mismatches like a small bias in the ANN or a soil water balance that is miscalibrated for the groundwater system as the effort required in identifying such issues far exceeded that available for this project.

The third general type of discrepancy is one in which modifications to a model to achieve a particular new scenario configuration are perceived by CalSim authorities (agency modelers and their consultants) as not being sufficiently consistent with some set of expectations assigned to that scenario. Ideally, the expectations by which the scenario is judged should align with the purpose for which that scenario was developed. This is not a guaranteed condition because expectations for CalSim results may differ widely across reviewers and applications. Given the potential for a mismatch in expectations assigned to CalSim 3 models, some discrepancies identified by CalSim authorities may primarily be a difference in philosophy with regard to application and interpretation of the CalSim 3 model by anyone who is not a CalSim authority – and thus resolution requires arbitration beyond a technical or mechanical modeling detail.

Other instances of divergence from expectation may have more direct causes and resolutions. These may arise in the CalSim inputs and range from simple errors in WRESL code (e.g. a switch to activate a functional flow requirement was not set to the active state for a scenario in which it should be active) to more complex issues such as multi-cycle accounting or tuning weights on a constraint. It is these types of errors and issues that a rapid but structured review process is best suited to catching and was the primary focus of the reviews performed by the COEQWAL modeling team.

Finally, discrepancies of this third type may arise due to an insufficiently robust modification of operational logic that either 1) fails to fully achieve the intended change in operation or 2) propagates undesired or unforeseen collateral changes to other parts of the model that were not intended to be modified. This last set of potential discrepancies is the most difficult to fully identify and characterize *a priori* and the most time intensive to correct after identification. COEQWAL modelers addressed these types of discrepancies to the extent possible allowed by the project timeline and according to the perceived severity of impact on the validity of the scenario in the context of the entire ensemble.

6. How the ensemble was produced and checked

6.1 Scenario development pipeline

COEQWAL scenarios were developed through a standardized process. Each scenario can be traced back through incremental changes to a single set of original source files published or provided by USBR or DWR. In some cases where a target scenario involves multiple major changes (land use and allocation priorities, for example), intermediate scenarios were created to check the incremental modifications (these intermediate scenarios are not included in the COEQWAL scenario library). Common demands and hydrology datasets were developed for hydroclimates and scenarios (i.e. reduced irrigated acreage and fallowed rice acreage) so that these components could be easily inserted when needed. To the extent practical within the project timeline, structural components of disparate models (i.e. USBR-sourced vs DWR-sourced models) were adjusted for consistency and uniformity, although some differences could not be reconciled in the time available (e.g. internal forecasting implementation in USBR models). Additionally, where possible, regularly changed operations, such as TUCP actions or minimum flow requirements, were implemented or utilized via switches (e.g. set a variable to 0, 1, 2 for different options) to simplify making these changes. Different hydroclimates were applied through standardized workflow that replaced existing time series input with the new target hydroclimate and set a new corresponding sea level rise condition. All post-processing and derived outputs (outcome levels, for example) were performed using a uniform workflow as well.

Despite these efforts, CalSim 3 is not structured such that all target operations can be implemented through a change to a single element without needing to modify or re-evaluate other components. More complex operational modifications required specific changes to WRESL code across many files that were not relevant to or necessary for any other scenarios. Examples of these include setting a minimum health and safety level of delivery for each M&I delivery end point in the model, modifying constraints or weights on a reservoir evaporation variable, or assigning a specific CVP allocation logic to increase Shasta carryover. See the Strategy briefs for an overview of the types of changes implemented in each scenario – some are unique to a scenario while others may be apparent across many scenarios.

6.2 Scenario-construction decisions under the project timeline

Producing an ensemble of this size and scope on the project timeline required making assumptions or simplifications that, in an ideal world, would be addressed through iterative refinement. This includes accepting a CalSim 3 implementation for a scenario targeted at an outcome simulated externally to CalSim 3 (e.g. winter run chinook salmon) before knowing whether the intended direction and magnitude of change was achieved. This is an unavoidable reality of applying a model as complex as CalSim 3 to analysis objectives as broad as those for COEQWAL on an ambitious timeline. However, the COEQWAL modeling team prioritized construction of scenarios that, within the context of the entire ensemble, provide variability across many different dimensions of broad interest to project partners while adhering to the conceptual goals of each scenario as best as possible.

6.3 Quality control applied to the ensemble

In-depth investigation of all 20,000+ time series generated from each scenario to ensure that no undesired or erroneous outcomes occurred anywhere in the model domain was not feasible. To limit the number of simple errors of the third type described above, the COEQWAL modeling team performed a standard set of checks on a selected subset of outputs from each simulation run. These checks included comparing monthly average (seasonality pattern) plots, exceedance plots, and direct time series comparison plots for storage, delivery, flow, and salinity variables. The review process was primarily a visual and qualitative check for reasonableness given the modifications made for a given scenario.

COEQWAL modelers also performed additional specific review of variables related to the intended outcome or modified component of each historical hydroclimate scenario. The goal with this review was to evaluate more specifically the degree to which the modified operations were achieving their intended purpose. Examples of this include comparing individual M&I delivery time series between the baseline and the two scenarios prioritizing M&I delivery levels to ensure delivery levels increased accordingly (if they were not already at or near the target). This stage of review did not conform to any universal set of evaluation rules as the sufficiency of each scenario was judged according to the type of modification being implemented, its relationship with other scenarios, and project timeline.

7. Review scope and what remains open

Review capacity across a library of this size is bounded, and this document does not claim that every value in every run has been inspected. Issues outside the coverage of those checks can remain in the raw output.

While systematic discrepancies present in all scenarios tend to cancel out when viewing results through a comparative lens (rather than absolute values), errors or discrepancies unique to a scenario can complicate interpretation. COEQWAL project results are presented first as summary metrics and figures intended to encourage comparison of patterns across scenarios. Small errors will likely not be legible in these summary comparison plots, but larger errors should be more obvious. The COEQWAL team has reviewed these pattern and inter-scenario comparisons for remaining issues, but there is a chance some persist. Discrepancies arising from smaller modeling errors, perceived insufficiency of model refinement, or simplified model assumptions may be present – the interested user should interrogate the raw model files and outputs for this purpose. The derivation of outcome levels is published so that the connection between raw outputs and the reported summary figures can be checked independently as well.

Any issues identified can be reported to the COEQWAL team via email at coeqwal@berkeley.edu. These issues will be logged and addressed as time and resources permit.