# Modeling actual flows evaluation

The Nature Conservancy submitted to the
Wildlife Conservation Board Streamflow Enhancement Program

Project ID 2022035

Jessica Ayers, Bronwen Stanford, Kirk Klausmeyer

October 27, 2025

# Introduction

To advance hydrologic predictions for streamflow management, Upstream Tech built a state-of-the-art set of models and datasets that utilize recent hydrological advancements in AI. In their approach, they used a type of neural network model called a Long Short-Term Memory (LSTM) model that is useful for time series data and learning relationships with input variables. Upstream Tech designed two models, the Unimpaired and Actual flows models, that leveraged data across the United States to predict actual and unimpaired flow in California. The Unimpaired model represents streamflow conditions in the absence of human alteration, while the Actual Flows model reflects real world conditions, whether unimpaired or influenced by human activities such as dam operations, water diversions and land use. The Unimpaired and Actual flows models share the same structure but differ in their inputs and how they incorporate upstream observations.

In this report, we evaluate the performance of the models Upstream developed for predicting daily streamflow for unimpaired and actual conditions. Our goal is to assess each model's ability to replicate observed flow patterns across California for a range of hydrologic conditions and watershed characteristics. We focus on examining how well the Unimpaired flows model performs for minimally disturbed watersheds that reflect streamflow in the absence of human alteration. We then evaluate model performance of the Actual model for all stream types, both minimally disturbed and disturbed. In addition, we also examined how the models performed at streamgages located downstream of dams and reservoirs using a dam impact index derived from the USGS. Finally, we evaluate model performance across a diverse range of hydroclimatic conditions using the regions defined by the ten Department of Water Resources (CDWR, 2023) and for different watershed sizes.

# Methodology

## Evaluation metrics

In this analysis, we evaluated model outputs from Upstream Tech's Unimpaired and Actual flows LSTM models comparing them with streamflow observations. We paired the daily modeled flow data from water years 2002-2022 with their corresponding USGS streamflow gages over the same period. In total, 328 USGS gages within California were used for training and evaluation of the Actual and Unimpaired models, with additional training gages located outside of California. For the Actual flows model, 230 California gages were used for training and 98 for evaluation. For the Unimpaired model, 41 and 54 were used for training and evaluation, respectively. In this report, we use the evaluation gages only so that we can evaluate how well the models perform in novel or ungaged locations.

We evaluated the performance of predictions based on the Nash-Sutcliffe Efficiency (NSE), mean coefficient of determination ($R^2$) and the percent bias (PBIAS). For $R^2$, a value of 1 indicates the model explains 100% of the variability in observed flow, and zero indicates the model explains none of the variability.

The Nash-Sutcliffe Efficiency evaluates the predictive performance of a model with the following equation:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{o,t} - Q_{s,t})^2}{\sum_{t=1}^{n}(Q_{o,t} - \bar{Q}_o)^2}$$

where $Q_{o,t}$ is the observed streamflow at time step t; $Q_{s,t}$ is the simulated or predicted streamflow at timestep t; $\bar{Q}_o$ is the mean of observed streamflow across all time steps; and n is the total number of observations. If the NSE is 1, it indicates a perfect model fit. A zero indicates that predictions are as accurate as the mean of observed data, while negative values indicate model predications are worse than the mean of observations.

The Percent Bias calculates the percentage relative differences using the following equation:

$$PBIAS = 100 \text{ X } \frac{\sum_{t=1}^{n} Q_{s,t} - Q_{o,t}}{\sum_{t=1}^{n} Q_{o,t}}$$

where $Q_{o,t}$ is the observed streamflow at time step t; $Q_{s,t}$ is the simulated or predicted streamflow at timestep t; and n is the total number of observations. If the PBIAS is zero, it indicates the model is in perfect agreement. A positive value indicates model overestimation, and a negative value indicates model underestimation.

When evaluating LSTM models for streamflow predictions, we focus on the NSE, as it evaluates how well the model's predictions match observed data and is sensitive to variability in the predicted time series. We first analyzed streamgages across the state to understand the variability of model performance for each model across California (Figure 1 and Table 1). We stratified streamgages using the USGS's characterization of minimally disturbed and disturbed watersheds (Table 2), and compared the unimpaired model only to minimally disturbed gages. The categories separate streamgages based on land use, dam presence and diversion records to define streams that experience alteration from human impacts (disturbed) versus those that are closer to natural streamflow (minimally disturbed).

We also used a dam impact index to extend our determination of how the actual flows model performed for gages below dams. We used data based on the USGS Dam impact and disturbance metrics for the conterminous U.S. dataset (Wieczorek et al., 2021) and the post-reservoir construction periods for monotonic trend analysis at streamgages in the U.S. dataset that quantitatively measure the impact of reservoir storage (Table 3). From Wieczorek et al. 2021, we used the EROM metric that estimates reservoir storage calculated as the number of days streamflow could be sustained based on the upstream reservoir storage. Headman and Hecht (2024) developed a decadal time series of cumulative normal reservoir storage capacity upstream of USGS streamgages. In this analysis, we used the most recent time period (2000 to 2020) to be consistent with the modeled data (2001-2022). When we compared the different datasets, we found the results to be similar. Thus, in this document we only report the results based on Headman and Hecht et al. (2024) for simplicity.

To understand variability in performance across a range of watershed characteristics, we stratified performance by watershed drainage area (Table 4). Finally, to understand how well

models performed in different regions of California and different hydroclimate regions, we grouped streamgages by the CDWR hydroclimate regions (Figure 2 and Table 5).

### Alteration assessment

As part of this assessment, we wanted to understand how well the predictions from the Actual flows model identified streamflow alteration across California. To evaluate hydrologic alteration, the Functional Flows Calculator (flowcalculator.codefornature.org) calculates functional flow metrics (FFM), which represent components of the hydrograph that are essential for ecological needs (Stein et al., 2021; Yarnell et al., 2020). For each season, the magnitude, timing, duration and rate of change are calculated from the input daily streamflow time series. In this analysis, we used the time series from the Actual flows model as input into the calculator to obtain predicted actual FFMs for each COMID (NHD medium resolution stream reach). We already have a validated dataset of predicted unimpaired FFMs for every reach in the state (Grantham et al. 2022). We then compared the predicted actual FFMs from the daily flows model against these existing predicted unimpaired FFMs to determine how well the model reproduced patterns of streamflow alteration across the state (Figures S1-4).

For each COMID and metric, the comparison yielded a classification of likely altered, likely unaltered, or indeterminate using the percentile-based criteria according to CEFWG appendix J (CEFWG 2021). For each metric, the alteration assessment first checks whether the median calculated (actual flows) value is within the 10th- 90th percentile modeled range and then assesses whether at least 50% of the calculated values fall within this 10th-90th percentile modeled range. If the first criterion is not met, the metric is considered altered. If the first criterion is met but the second is not, it is considered indeterminate. For each functional flow component, we determined a COMID as likely altered if at least one metric within that component was classified as likely altered. If none were classified as likely altered and there was at least one metric with an indeterminate classification, then that component was classified as indeterminate. All other COMIDs were classified as likely unaltered.

## Results and discussion

### Statewide performance

On average, there was poor performance of the Unimpaired and Actual flows models at the statewide scale, with large variability in performance (Figure 1 and Table 1). The mean NSE values for the Unimpaired and Actual Flows models were -0.40 and -1.35, respectively. Overall, these negative values reflect poor model fit as the predictions are worse than if we used the mean of observed streamflow as the prediction; however, it's important to note that that model did not have access to any observed streamflow measurements for these locations.

The median value for the NSE was 0.26 for both models, highlighting that the averages were skewed by the outliers and models may perform better in some locations, as highlighted by the spatial distribution in Figure 1. The mean $R^2$ values were 0.45 and 0.44 for the Unimpaired and Actual Flows models, respectively. Finally, the PBIAS for the Unimpaired and Actual flows model was -22.62 and -21.11, respectively. The negative PBIAS values show that the models underestimated streamflow in most cases. The results suggest that while the models were able to

capture some aspects of flow variability and seasonal patterns, they did not accurately reproduce streamflow magnitudes.

Across the state, performance varied widely, with both models performing better in northern California, the Sacramento Valley and parts of coastal California. Poorer performance was observed in southern California and the Sierra. Along the south coast, both models had difficulty predicting flows (NSE between 0 and 0.5), and most basins with high negative bias were observed in more mountainous regions, indicating the model may struggle with watersheds that have complex geology. The models performed better in wetter climates where more perennial streams are prevalent, compared to more arid regions of California with intermittent streams. Additionally, these results suggest that performance could be better in areas with less alteration (i.e., northern California) and worse in highly altered and/or groundwater dependent basins in the south.

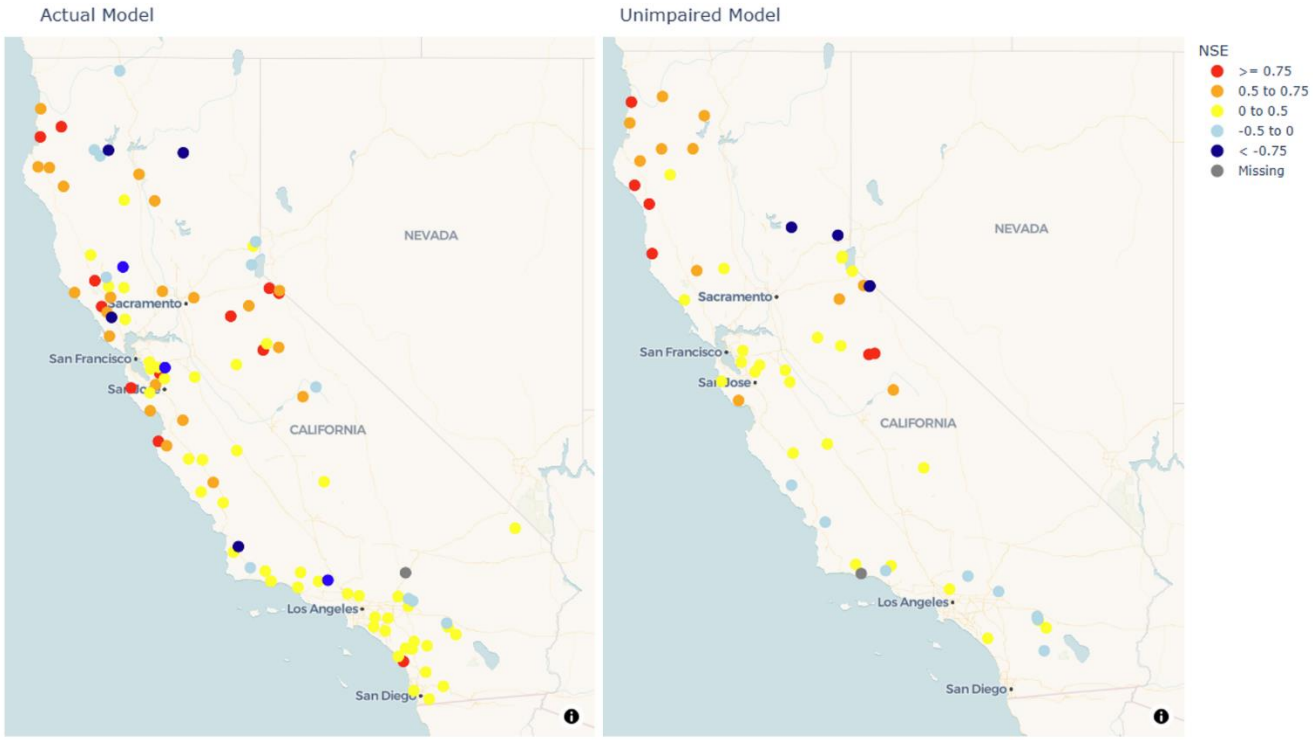



**Figure 1**: Maps displaying the NSE values for daily flows at the evaluation 98 and 54 USGS streamgages for the Actual and Unimpaired model, respectively.

**Table 1**: Mean performance statistics for the Unimpaired and Actual Flows models across the evaluation USGS streamgage dataset in California. The performance metrics include the $R^2$, NSE and PBIAS.

| Model | N | $R^2$ | NSE | PBIAS |
|---|---|---|---|---|
| Unimpaired | 54 | 0.45 | -0.40 | -22.62 |
| Actual | 98 | 0.44 | -1.35 | -21.11 |

**Disturbed and minimally disturbed gages and dam impacted streams**

Minimally disturbed and disturbed streamgages were extracted to understand the difference in performance between the two models across the state (Table 2). For the Unimpaired model, we only compared predicated daily streamflow values against minimally disturbed evaluation gages, as the model predicts unimpaired streamflow conditions. Here the $R^2$, NSE, and PBIAS were 0.48, 0.08, and -31.36, respectively. For the Actual flows model, we used both minimally disturbed and disturbed evaluation gages, as the model should represent flows in all streams. For minimally disturbed gages, the $R^2$, NSE, and PBIAS were 0.43, -7.96, and –25.66, respectively. For the disturbed gages with the Actual flows model, the $R^2$, NSE and PBIAS had values of 0.45,0.22, and –18.44, respectively.

Overall, the statistical results show key differences in the performance of each model between different types of land use and disturbance. For minimally disturbed gages, the Unimpaired model performed better than the Actual flows model. However, the Actual and Unimpaired models both had negative NSE and PBIAS values for minimally disturbed gages, which indicates that both models underestimated streamflow in basins with natural conditions. For disturbed gages, the Actual Flows model performed similarly to minimally disturbed gages, indicating that the models may represent some aspects of altered and unimpaired streamflow.

**Table 2**: Summary of model performance for the Actual and Unimpaired flows models for USGS categories of minimally disturbed and disturbed watersheds using the evaluation gage dataset. The statistics include the number of streamgages in each category (N), and the $R^2$, NSE and PBIAS values, which represent the mean value for all streamgages in each category.

| | **Actual** | | | | **Unimpaired** | | | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **$R^2$** | **NSE** | **PBIAS** | **N** | **$R^2$** | **NSE** | **PBIAS** |
| Minimally disturbed | 29 | 0.43 | -7.96 | -25.66 | 52 | 0.48 | 0.08 | -31.36 |
| Disturbed | 69 | 0.45 | 0.22 | -18.44 | | | | |

We also compared the Unimpaired and Actual Flows model performance for USGS streamgages below dams and reservoirs (impacted) and those with little or no impairment from dam regulation (nonimpacted; Table 3). Of the evaluation gages used for the Actual flows model, there were 39 gages with impacted and 28 gages nonimpacted conditions using data available from the Headman and Hecht (2024) dataset (not all evaluation gages were included in the dataset). The Actual flows model performed better at impacted gages ($R^2$ =0.50; NSE=0.28, PBIAS=-0.88) than nonimpacted gages ($R^2$ =0.50; NSE= -8.29, PBIAS=-4.58), which suggests that reservoir releases were potentially more accurately represented by the Actual flows model than natural conditions or smaller diversions.

**Table 3**: Summary of model performance for the Actual and Unimpaired flows models for categories of impacted and nonimpacted streams with dam operations and reservoirs (Headman and Hecht 2024). The statistics include the number of streamgages in each category (N), and the $R^2$, NSE and PBIAS values, which represent the mean value for all streamgages in each category.

| | **Actual** | | | | **Unimpaired** | | | |
|---|---|---|---|---|---|---|---|---|
| | **N** | **$R^2$** | **NSE** | **PBIAS** | N | **$R^2$** | **NSE** | **PBIAS** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Nonimpacted | 28 | 0.50 | -8.29 | -4.58 | 37 | 0.54 | 0.26 | -32.15 |
| Impacted | 39 | 0.50 | 0.28 | -0.88 | | | | |

**Drainage area**

We segregated gages by drainage area to evaluate how well the models predicted streamflow across different watershed areas (Table 4). Overall, there were not clear patterns by drainage area. The unimpaired flows model showed low NSE for the smallest watersheds, suggesting that below 30 sq km the results may not be reliable. Otherwise, results were variable, suggesting that other factors are driving model performance.

**Table 4**: Model performance metrics for the Actual Flows and Unimpaired Flows models summarize for different drainage area categories using the evaluation gage dataset. The statistics include the number of streamgages in each category (N), and the $R^2$, NSE and PBIAS values.

| | | **Actual** | | | **Unimpaired** | | | |
|---|---|---|---|---|---|---|---|---|
| **Drainage Area** km$^2$ | **N** | $R^2$ **mean** | **NSE** mean | **PBIAS** mean | N | $R^2$ mean | **NSE** mean | **PBIAS** mean |
| <= 30 | **10** | **0.45** | 0.25 | -7.55 | 10 | 0.40 | -3.13 | 0.57 |
| 30-50 | 5 | 0.33 | 0.11 | -12.32 | 5 | 0.44 | 0.29 | -4.10 |
| 50-70 | 6 | 0.45 | 0.33 | -4.98 | 6 | 0.38 | 0.17 | -10.89 |
| 70-100 | 4 | 0.11 | 0.03 | -15.41 | 4 | 0.43 | 0.30 | 5.37 |
| 100-150 | 10 | 0.40 | -0.13 | -9.67 | 5 | 0.33 | 0.12 | 5.63 |
| 150-200 | 7 | 0.49 | 0.41 | -10.77 | 5 | 0.48 | 0.37 | -37.57 |
| 200-300 | 9 | 0.47 | 0.35 | -41.88 | 5 | 0.44 | 0.34 | -23.60 |
| 300-400 | 6 | 0.26 | -0.06 | -24.39 | 3 | 0.55 | 0.38 | -60.61 |
| 400-500 | 6 | 0.40 | -39.45 | -2.25 | 2 | 0.53 | 0.42 | -62.55 |
| 500-700 | 6 | 0.55 | 0.53 | -43.28 | 3 | 0.55 | 0.49 | -110.45 |
| 700-1000 | 2 | 0.71 | 0.60 | -13.92 | 4 | 0.80 | 0.72 | -17.67 |
| 1000-1500 | 9 | 0.53 | 0.41 | -15.67 | 0 | | | |
| 1500-2000 | 3 | 0.45 | -0.19 | 280.02 | 2 | 0.71 | 0.65 | -289.12 |
| Greater than 2000 | 15 | 0.56 | 0.31 | -75.87 | 0 | | | |

**Regional analysis**

Finally, we evaluated model performance by CDWR hydroclimate region (Figure 2 and Table 5). We found that performance varied substantially by region. The San Joaquin River region had good performance for both models, where the NSE values were 0.47 and 0.54 for the Actual and Unimpaired models, respectively. The North Coast region had the strongest results from the Unimpaired flows model (NSE = 0.60), but the Actual flows model performed poorer (NSE= 0.14). Next, the San Francisco Bay performed the best with NSE values of 0.36 and 0.22 for the Actual and Unimpaired models, respectively.

Moderate performance was observed in the North Lahontan, South Coast and Central Coast regions where NSE values ranged between 0.35 and 0.17 for the Actual flows model. However, for these regions, the Unimpaired flows model had consistently worse NSE values (between 0.19 and –2.70). Overall, the models perform well to decent along coastal California, although the model performance between the Actual and Unimpaired differed. Generally, we found that as sites move southward, the Unimpaired model exhibited weaker predictive skill, while performance for the Actual model was more variable.

Finally, models performed worse in the Colorado River, Tulare Lake, and Sacramento River, two basins with high hydrologic alteration. The NSE values for the Actual flows model were 0.01, 0.07, and –26.26 for the Colorado River, Tulare Lake, and Sacramento River, respectively. These river basins are heavily regulated by reservoir networks with complex water management. Both the Colorado River and Sacramento River have characteristically snowmelt headwaters with large-scale agricultural practices, including groundwater pumping, interbasin water transfers and surface water diversions that modify streamflow dynamics. Consequently, model performance was reduced in these heavily managed systems where human alteration is variable and low. However, performance was also poor for the Unimpaired model: the NSE values were –0.03, 0.04, and –2.65 for the Colorado River, Tulare Lake, and Sacramento River, respectively. Finally, we did not evaluate data in the South Lahontan because only 1 evaluation gage was located within the basin boundaries for both models.

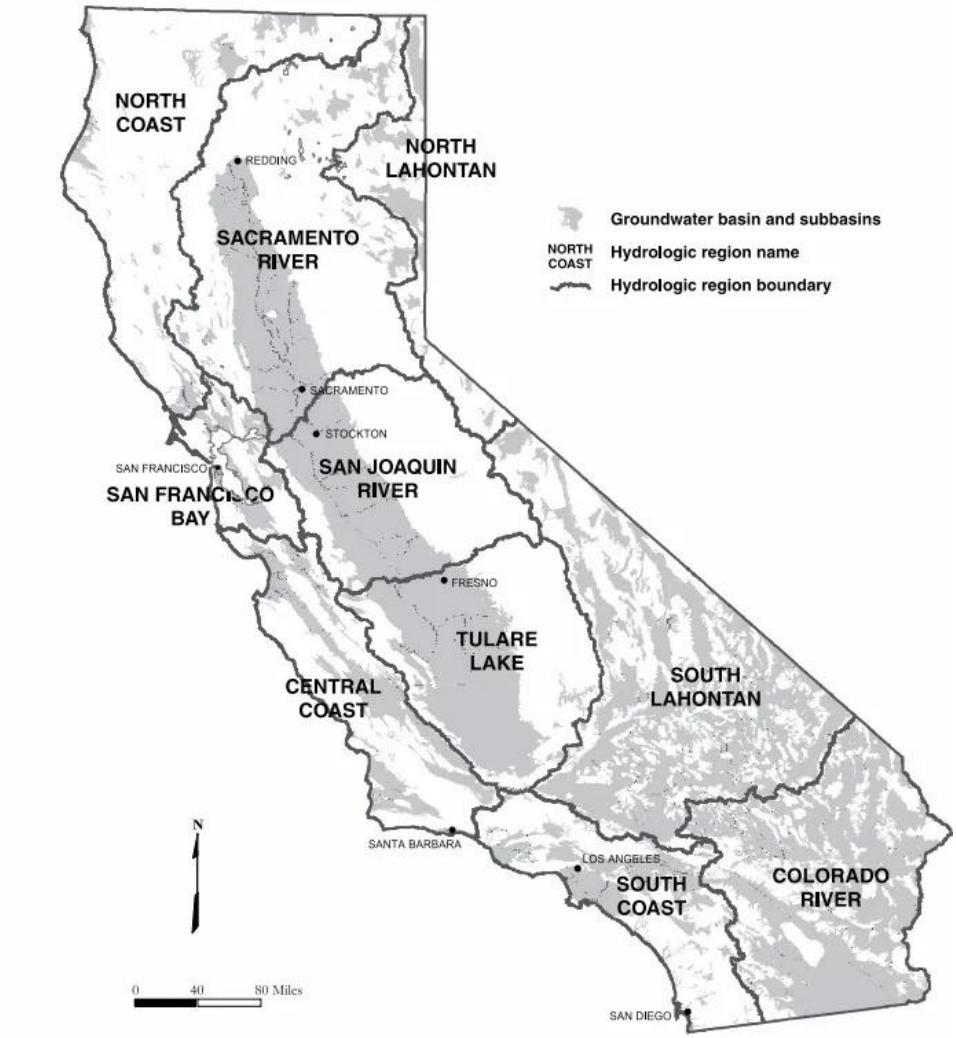



**Figure 2**: Map showing the CDWR hydrologic regions used to split model performance across the state. CDWR subdivided California into ten hydrologic regions that correspond to the State's major watershed drainage basins.

**Table 5**: Model performance metrics for the Actual Flows and Unimpaired Flows models summarized for ten CDWR hydrologic regions. The statistics include the number of streamgages in each category (N), and the $R^2$, NSE and PBIAS values. This gage set includes the evaluation gages only.

| | | Actual | | | | Unimpaired | | |
|---|---|---|---|---|---|---|---|---|
| **Region** | N | $R^2$ mean | NSE mean | PBIAS mean | N | $R^2$ mean | NSE mean | PBIAS mean |
| North Coast | 18 | 0.56 | 0.14 | -59.29 | 12 | 0.70 | 0.60 | -88.47 |
| San Joaquin River | 9 | 0.64 | 0.47 | 48.85 | 8 | 0.63 | 0.54 | -14.00 |
| San Francisco Bay | 11 | 0.52 | 0.36 | -8.71 | 5 | 0.37 | 0.22 | -6.62 |
| Central Coast | 14 | 0.40 | 0.25 | -6.27 | 5 | 0.31 | 0.19 | -32.09 |
| South Coast | 24 | 0.34 | 0.17 | -6.81 | 5 | 0.34 | 0.14 | -8.17 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| North Lahontan | 6 | 0.54 | 0.35 | -27.41 | 7 | 0.55 | -2.70 | -21.57 |
| Tulare Lake | 2 | 0.33 | 0.07 | -7.52 | 2 | 0.20 | 0.04 | -6.76 |
| South Lahontan | 0 | | | | 1 | 0.28 | -0.11 | -12.87 |
| Colorado River | 4 | 0.08 | 0.01 | -0.83 | 5 | 0.11 | -0.03 | -1.88 |
| Sacramento River | 9 | 0.38 | -26.26 | -46.12 | 3 | 0.48 | -2.65 | -9.75 |

**Alteration assessment**

Although the CEFF alteration assessment is a powerful tool to determine streamflow impairment, the issues with model performance limited the usefulness of the analysis. To perform this assessment, we compared the results of one statewide model (the Actual flows model) against another statewide model (the existing natural flows predictions), which introduces a lot of potential for error. When we evaluated the results of the assessment (Figures S1-4), the statewide patterns of seasonal alteration did not represent realistic patterns of alteration. For example, we would expect to see more frequent "likely unaltered" classifications in the headwaters and tributaries and "likely altered" classifications in the mainstem rivers. However, the maps produced failed to replicate aspects of this pattern, and in some watersheds the results had the opposite pattern. Additionally, in the spring, nearly every COMID in the state was classified as likely altered (Table 6, Figure 3). Although we might expect a high degree of alteration in the spring metrics in regions with changing snowmelt patterns and human alteration, the uniform alteration classification likely does not reflect real world conditions.

**Table 6**: Alteration assessment results summarized by season. If any metric is altered for a COMID, the entire functional flow component is considered altered, so "total likely altered" is the sum of the number of reaches with 1, 2, 3, and 4 metrics altered. Some metrics were not able to be calculated in enough years to perform an accurate alteration assessment. These are noted in the "Missing data" row. A total of 107,910 reaches were considered for this analysis.

| | Dry season | Fall pulse | Spring recession | Wet season |
|---|---|---|---|---|
| 1 altered metric | 27% | 9% | 45% | 20% |
| 2 altered metrics | 14% | 1% | 23% | 16% |
| 3 altered metrics | 4% | 0% | 5% | 2% |
| 4 altered metrics | 2% | -- | 1% | 1% |
| **TOTAL likely altered** | **48%** | **11%** | **73%** | **39%** |
| | | | | |
| **Indeterminate** | **3%** | **2%** | **3%** | **9%** |
| **Total likely unaltered** | **41%** | **18%** | **8%** | **41%** |
| *Missing data* | *8%* | *69%* | *16%* | *12%* |

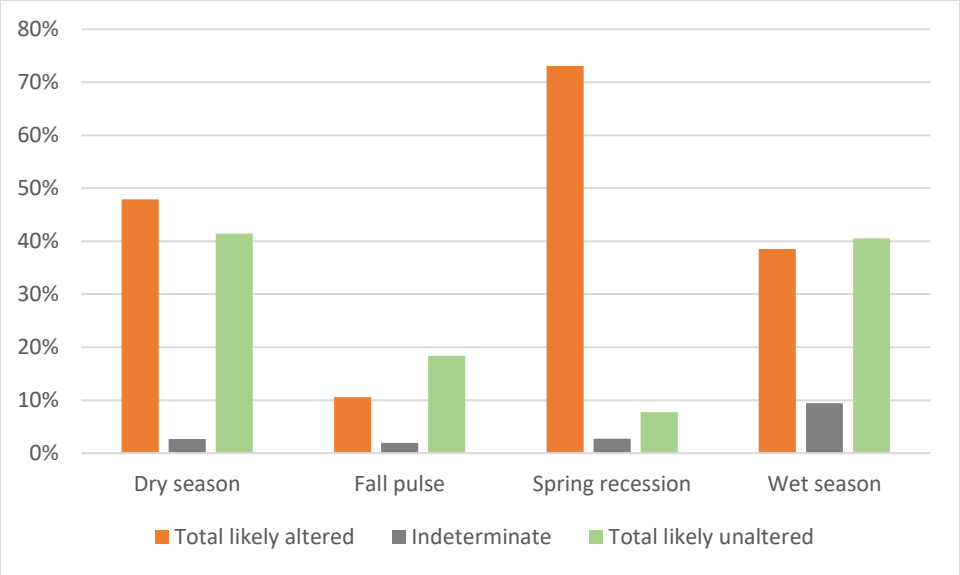



**Figure 3**: Plot showing the percent of stream reaches in California falling into each of the three alteration categories by functional flow component. These results do not appear to be realistic.

As discussed previously, the performance of the Actual flows model was not reliable, especially in regions with complex hydrologic regimes (e.g., inland and southern California). Since the alteration assessment depends directly on the accuracy of the daily predictions, its outputs inherently have uncertainties. As a result, the statewide alteration results should be interpreted with extreme caution and/or used only as a screening tool rather than a publishable analysis. As a result, we have chosen to focus on the more interpretable daily flows, which can more readily be compared to observed data.

## Conclusions

In this report, we evaluated the performance of the Unimpaired and Actual flows models developed by Upstream Tech to predict daily streamflow in California. We evaluated performance at 328 USGS streamgages for a variety of hydrologic conditions and watershed characteristics (e.g., drainage area, by region, and disturbance level). Overall, we found that the Unimpaired model performed better in northern and coastal regions of California, and both models performed worse in southern inland regions and the Sierra. In northern California, there are more perennial streams that the models predicted more accurately than the flashy and intermittent streams in the south. Additionally, streams in the north have less disturbance (e.g., urbanization and agricultural practices) compared to southern regions and the Sacramento River. The Actual flows model performed equally well at predicting flows below reservoirs and not, and had higher performance in disturbed than least-disturbed basins, which suggests that it is able to represent at least some aspects of streamflow alteration. In least-disturbed basins, the Unimpaired flows model performance was higher.

Overall, both models tended to underestimate streamflow, as was seen by the consistently negative PBIAS values. Inspection of individual hydrographs suggests that commonly some higher flow events are either missed or have lower than observed magnitude. Model accuracy for the

Unimpaired model decreased in streams with small drainage areas (i.e., less than 30 $km^2$). Small catchments often have flashy hydrologic responses, which can be difficult to capture because they usually depend on small rainfall events, especially at the daily time step. In addition, small basins are usually more sensitive to local disturbances that are not well represented in course catchment variables that are input into the models.

The next step for this project will be posting a final dataset onto a TNC website for access by the public. To support responsible use of the modeled streamflow data on the TNC website, we will publish a user report that clearly outlines the data's strengths, limitations and appropriate applications. In practice, individuals applying these predictions should evaluate their reliability for the specific management and conservation needs of their study. At this time, we feel this dataset is most useful for research and modeling applications, rather than informing management decisions on any individual river.

**References**


California Department of Water Resources. (2023). *i03 Hydrologic Regions* [Dataset]. California Open Data. Retrieved *[September 2025]*, from https://data.ca.gov/dataset/i03-hydrologic-regions

California Environmental Flows Working Group (CEFWG). 2021. California Environmental Flows Framework Version 1.0. California Water Quality Monitoring Council Technical Report 65 pp. https://ceff.ucdavis.edu/tech-report

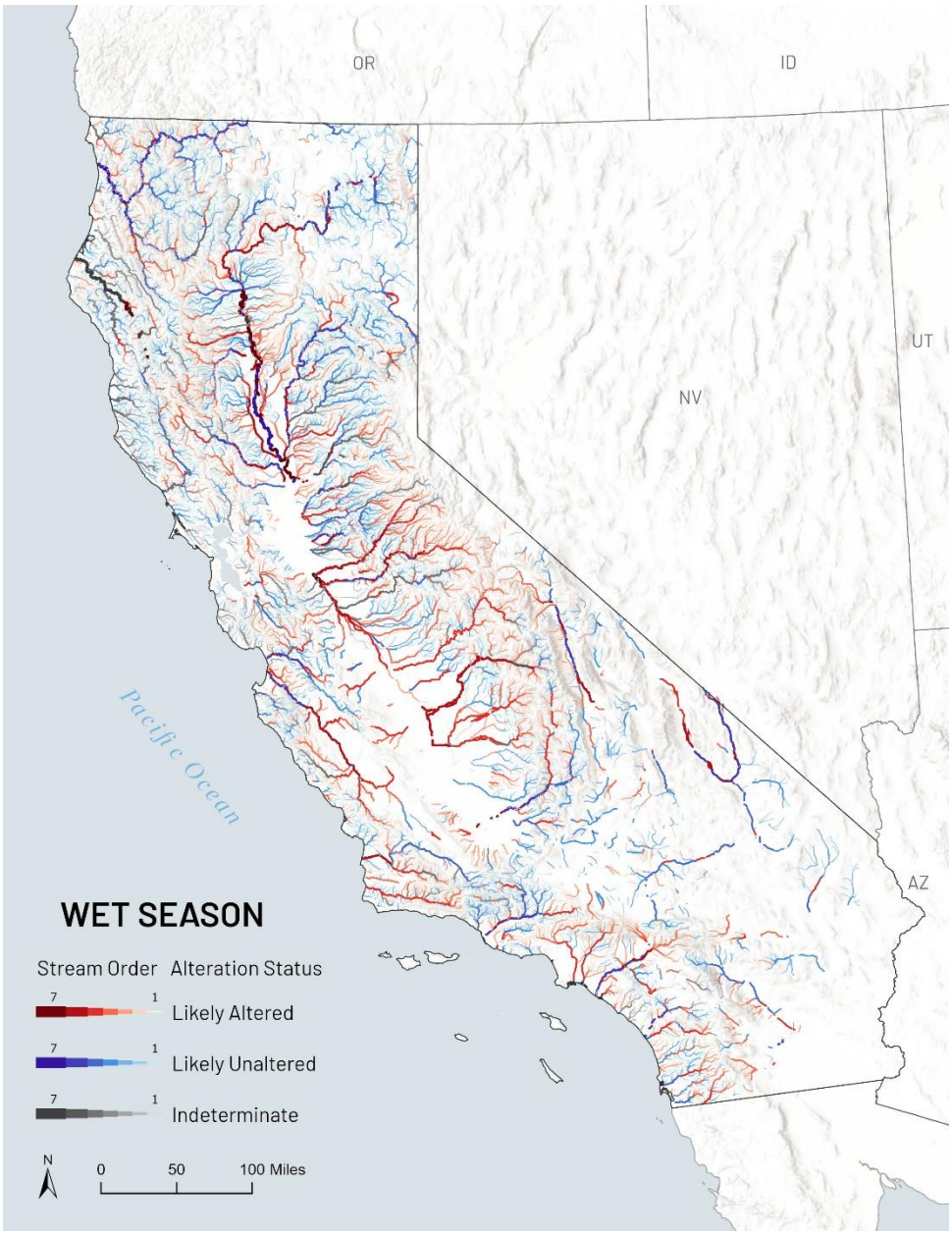
Headman, A.O. and Hecht, J.S. 2024. Identifying post-reservoir construction periods for monotonic trend analysis at streamgages in the United States (ver. 2.1, May 2025): U.S. Geological Survey data release, https://doi.org/10.5066/P922P61Z.

Stein, E.D., J. Zimmerman, S.M. Yarnell, B. Stanford, B. Lane, K.T. Taniguchi-Quan, A. Obester, T.E. Grantham, R.A. Lusardi, and S. Sandoval-Solis. (2021). The California Environmental Flows Framework: Meeting the Challenges of Developing a Large-Scale Environmental Flows Program. Frontiers in Environmental Science 9:769943. doi: 10.3389/fenvs.2021.769943

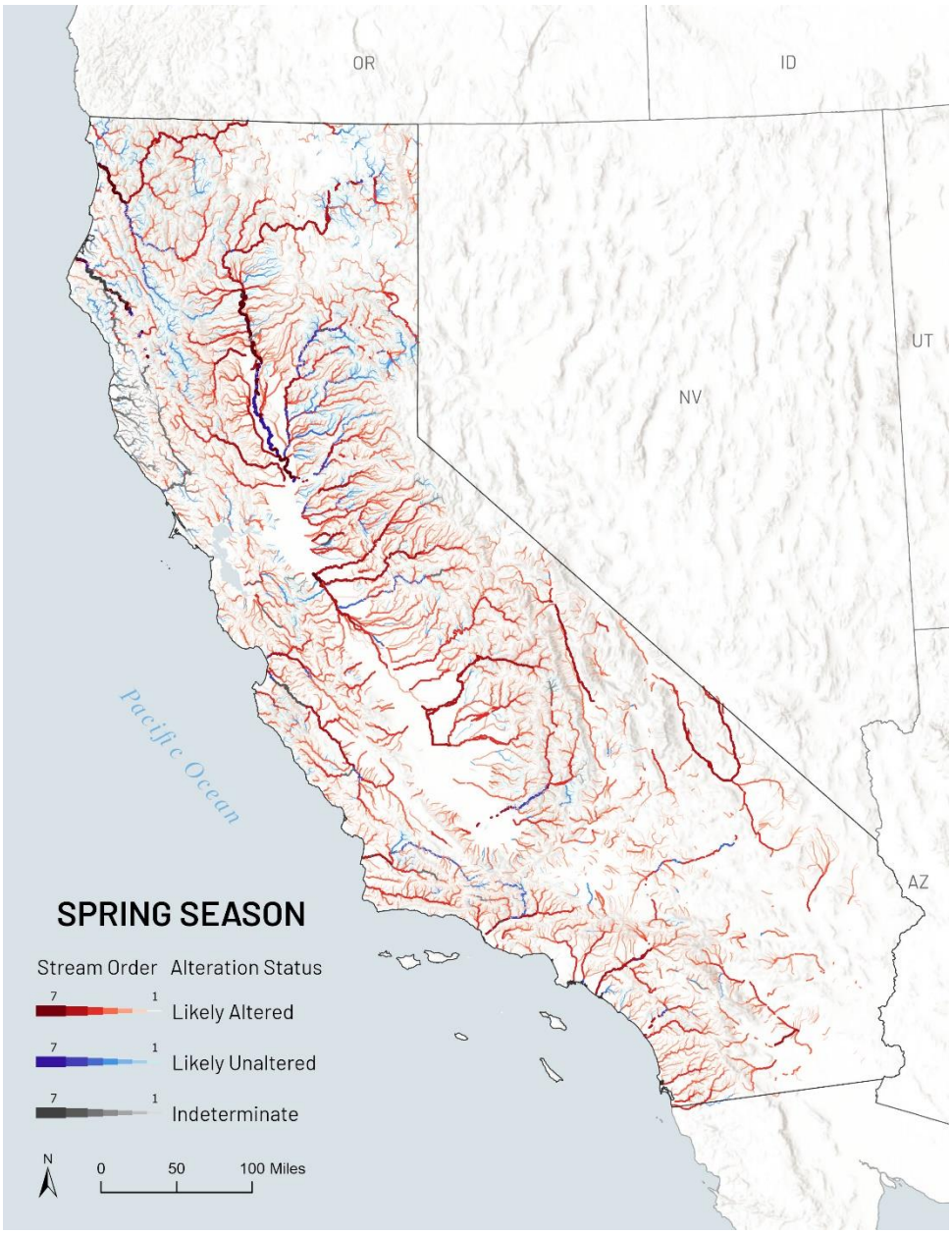
Wieczorek, M.E., Wolock, D.M., and McCarthy, P.M., 2021, Dam impact/disturbance metrics for the conterminous United States, 1800 to 2018: U.S. Geological Survey data release, https://doi.org/10.5066/P92S9ZX6.

Yarnell, S.M., E.D. Stein, J.A. Webb, T.E. Grantham, R.A. Lusardi, J. Zimmerman, R.A. Peek, B.A. Lane, J. Howard, and S. Sandoval-Solis. (2020). A functional flows approach to selecting ecologically relevant flow metrics for environmental flow applications. River Research and Applications, 36(2), 318–324. doi: 10.1002/rra.3575

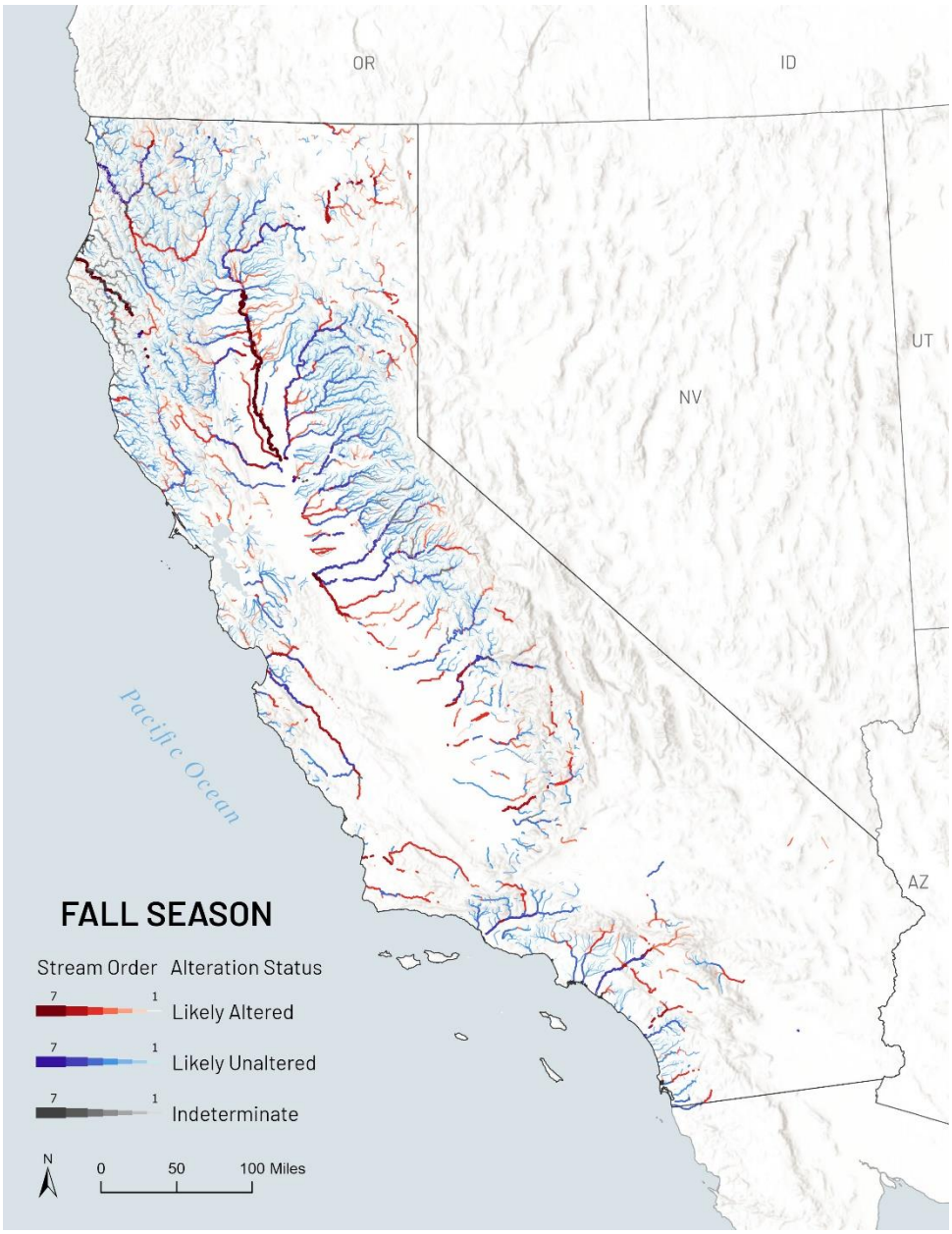
**Supplemental Figures**



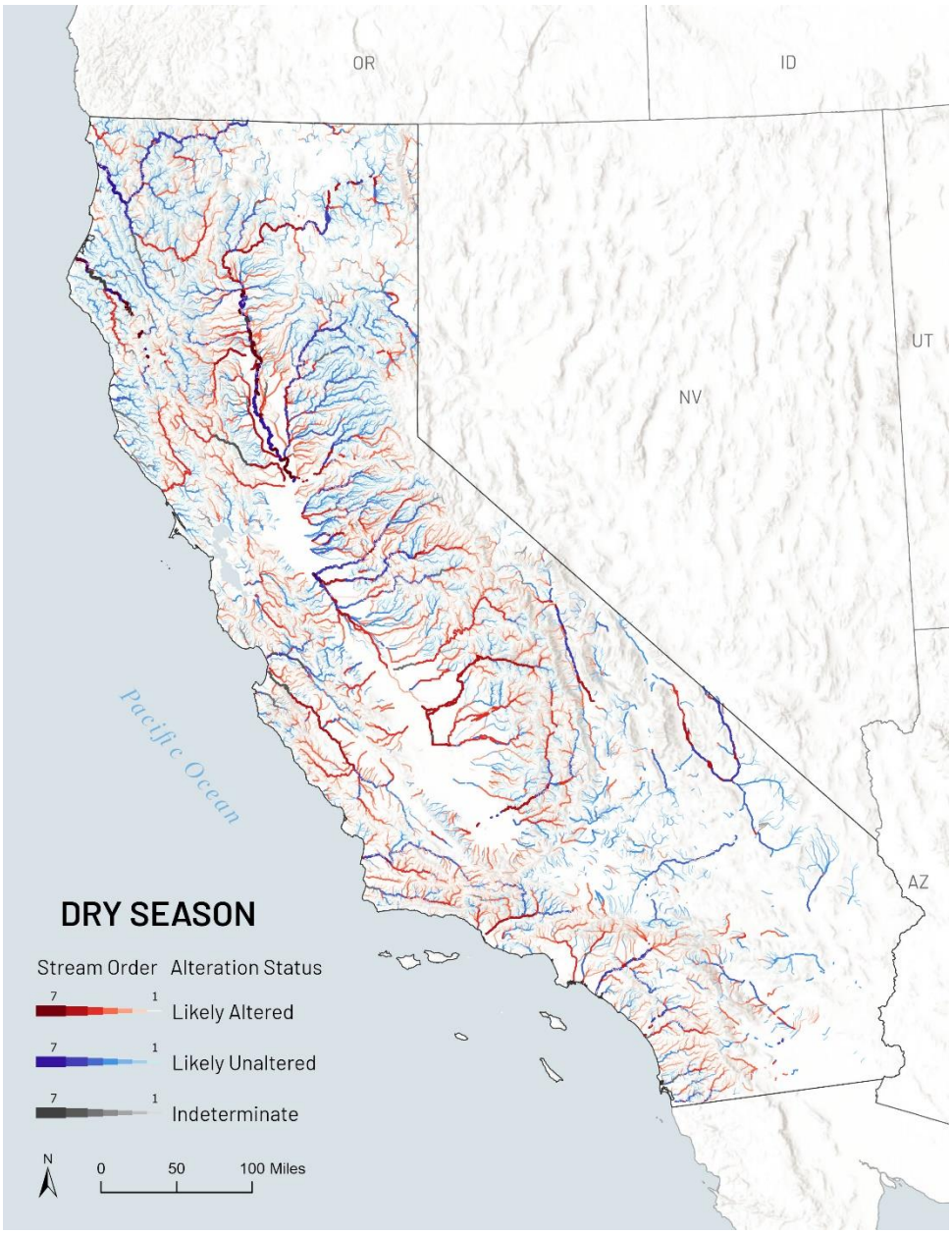

**Figures S1:** Map of wet season baseflow hydrologic alteration for all COMIDs in California based on the functional flow metrics calculated from the Actual flows model daily streamflow time series. Red indicates likely altered where at least one metric within the flow component (timing, duration, magnitude) was identified as altered. If none were likely altered, but at least one metric within the flow component was identified was indeterminate, the color grey indicates that alteration status is indeterminate. All other COMIDs were classified as likely unaltered. Width of the line indicates stream order.




**Figures S2:** Map of spring recession flow alteration for all COMIDs in California based on the functional flow metrics calculated from the Actual flows model daily streamflow time series. Red indicates likely altered where at least one metric within the flow component (timing, duration, magnitude, rate of change) was identified as altered. If none were likely altered, but at least one metric within the flow component was identified was indeterminate, the color grey indicates that alteration status is indeterminate. All other COMIDs were classified as likely unaltered. Width of the line indicates stream order.




**Figures S3:** Map of fall pulse flow hydrologic alteration for all COMIDs in California based on the functional flow metrics calculated from the Actual flows model daily streamflow time series. Red indicates likely altered where at least one metric within the flow component (timing, duration, magnitude) was identified as altered. If none were likely altered, but at least one metric within the flow component was identified was indeterminate, the color grey indicates that alteration status is indeterminate. Width of the line indicates stream order. Note: the fall pulse flow does not occur in all streams and years.




**Figures S4:** Map of dry season baseflow alteration for all COMIDs in California based on the functional flow metrics calculated from the Actual flows model daily streamflow time series. Red indicates likely altered where at least one metric within the flow component (timing, duration, magnitude) was identified as altered. If none were likely altered, but at least one metric within the flow component was identified was indeterminate, the color grey indicates that alteration status is indeterminate. All other COMIDs were classified as likely unaltered. Width of the line indicates stream order.