DRAFT | Peer Review Purposes Only | Not for Citation

Overview of Monitoring Data Availability and Processing for the Spring-Run Juvenile Production Estimate Models

Authors

Ashley Vizek, FlowWest; Liz Stebbins, FlowWest; Erin Cain, FlowWest; Brett Harvey, California Department of Water Resources

Acknowledgments

The data used in this modeling effort are rich and extensive—in some cases, data collection has been ongoing since the late 1990s. This work would not have been possible without the field staff that collected daily rotary screw trap data over many years, and the data stewards who managed that data. We specifically acknowledge the current data collectors and stewards Anna Allison, Nicolas Bauer, Drew Huneycutt, Jeanine Phillips, and Corey Fernandez who assisted in this effort, curating and making compatible data used in analysis, and providing valuable insights that contribute to the modeling process.

Work by FlowWest on this project was supported by the California Department of Water Resources.

Contents

Tables

Figures

Acronyms and Abbreviations

API	application programming interface
BT-SPAS	Bayesian Temporally Stratified Population Analysis System
CDFW	California Department of Fish and Wildlife
DWR	California Department of Water Resources
EDI	Environmental Data Initiative
HFC	high-flow channel
JPE	juvenile production estimate
LFC	low-flow channel
PLAD	probabilistic length-at-date
RST	rotary screw trap
USFWS	U.S. Fish and Wildlife Services

1 Introduction

This chapter provides an overview of available monitoring data and the process for updating datasets for models to forecast the annual production of Central Valley spring-run Chinook salmon (*Oncorhynchus tshawytscha*) (spring-run) juvenile outmigrants, also known as the spring-run juvenile production estimate (JPE). For those who want to take a deeper dive, links are embedded throughout this chapter, and are listed in Section 4. These links offer more detailed information about monitoring, data types, and data processing to ready datasets for modeling.

The monitoring sites and data types produced at those sites are summarized in Figures 1, 2, and 3. The Stream Teams collect these data; Table 1 lists the monitoring program leads and Data Stewards for each monitoring program. Table 2 summarizes the general types of data and the general spring-run JPE model types that use them. Table 3 lists additional details by data collection location.

Compiled and processed datasets used in spring-run JPE modeling are available in the SRJPEdata R package on GitHub. Table 2 also highlights key datasets (Vizek et al. 2025) and offers more detailed reference documentation for all datasets included in the SRJPEdata R package, along with more detailed documentation about data processing, which are available on GitHub via the SRJPE subdomain's Articles tab.

Preparing data for use in spring-run JPE modeling required making decisions about what data to use. Geographic scope was defined in the 2021 *Interim Monitoring Plan for the Spring-Run Chinook Salmon Juvenile Production Estimate Science Program* (Interim Monitoring Plan) (Allison et al. 2021), which also describes data collection methods.

2 Summary of Data Types and Data Processing

2.1 Data Types

Rotary screw trap (RST) monitoring data from spring-run JPE streams was used for juvenile outmigrant abundance modeling. Some streams operate RSTs at multiple sites (Vizek et al. 2025). All RST data from tributary sites were combined for use in the BT-SPAS-X models because the hierarchical structure of the model leverages data across sites to reduce uncertainty in model output for individual sites. However, individual output from all sites was not used in the stock recruit or in-season outmigrant models.

As part of the spring-run JPE program, two new mainstem RST monitoring sites were initiated: one on the lower Sacramento River near the point of Delta entry, and the other on the Lower Feather River (Figure 2). These datasets do not yet include enough years for use in spring-run JPE modeling. Once these programs accrue enough years of data, they will likely be considered for use in updated spring-run JPE modeling. Genetic samples are collected at all currently-operated RSTs that supply data for the spring-run JPE, including the new mainstem sites. More detailed information about RST sites can be found on the Articles tab on SRJPEdata website.

Stock recruit models required adult data from the spring-run JPE streams. Four types of adult data are collected and were tested in stock recruit models:

- Video passage counts
- Holding (over-summering) counts
- Carcass estimates
- Redd counts

More detailed information about data collection methods and the rationale for decisions about which adult data type were suitable for stock recruit modeling is available on the SRJPEdata website. Detailed information about the rationale for selecting the appropriate RST site for generating a response variable for stock recruit modeling is available on GitHub on the Articles tab on SRJPEdata website. These are the same sites that were used to produce juvenile abundance estimates from BT-SPAS-X for in-season outmigrant modeling.

The Data Management Team worked with the Modeling Team, Stream Teams, and regional subject matter experts to develop datasets of physical covariates that could be tested as predictor covariates for the various spring-run JPE models. These covariates were mainly based on flow and temperature monitoring data. Physical

covariate testing for specific models is described in the chapters for those models. Additional information on which gages were used to provide source data is available on the Environmental Gage Data page. Processing of physical data to develop covariates for models is described in Section 2.2.5.

2.2 Making Data Useful for Spring-Run Juvenile Production Estimate Modeling

Data used for the spring-run JPE were and continue to be collected by different monitoring programs run by multiple agencies, are collected in many different locations using diverse methods, and using a range of different reporting standards and data management platforms. To be useful for the spring-run JPE, the Stream Teams and Data Management Team collaborated on five phases of curation and processing:

1. Data Discovery and Outreach
2. Quality Control and Exploration
3. Data Publication
4. Standardization and Integration
5. Data Processing to Develop Model Input Parameters

These phases are described in the following sections.

2.2.1 Phase 1, Data Discovery and Outreach

In Phase 1, the Stream Teams and Data Management Team began data discovery for the spring-run JPE in 2021 based on data identified in the *Incidental Take Permit Spring-Run Chinook Salmon Juvenile Production Estimate Science Plan 2020–2024* (Science Plan) (DWR et al. 2020) and the Monitoring Plan. At that time, the majority of data were unpublished or were only available in PDF reports. Throughout this phase, the Data Management Team captured important metadata including methods, change in methods over time, site characteristics, data limitations, and other information, and made this information available and usable.

2.2.2 Phase 2, Quality Control and Exploration

Following the acquisition of raw machine-readable data (typically in Microsoft Excel or Access data format), R Markdowns were developed that included data visualizations for exploration and quality control of each variable included in the datasets. Work during Phase 2 also captured the differences in data structure, methods, and naming conventions in datasets across the different monitoring programs and data types. This information was critical to developing a standard schema to integrate the data.

2.2.3 Phase 3, Data Publication

In Phase 3, the Data Management Team began working with each of the Stream Teams to publish their historical RST data on the Environmental Data Initiative (EDI) and continues to work with the Stream Teams to publish historical adult data and new data as they are collected. EDI was selected as the best available publication site based on a number of factors including public accessibility and documentation standards (Harvey et al. 2022). RST data are updated biweekly and adult data are updated annually. Datasets can be found easily by going to the EDI website (edirepository.org) and search for "SRJPE" in the "Find Data" page.

2.2.4 Phase 4, Standardization and Integration

In Phase 4, the Data Management Team worked with the Stream Teams to develop standard data schemas for each data type to integrate data from different monitoring programs so they could be used together in spring-run JPE models. This information can be found on GitHub at SRJPE in the "system-design-docs" area. Data are pulled from EDI via an application programming interface (API), wrangled into a standard format, and made available in an R data package called SRJPEdata. SRJPEdata includes documentation for each dataset available in the package (go to the Reference area and select the dataset of interest) as well as articles describing data processing (click the Reference area's Articles tab for a drop-down list).

2.2.5 Phase 5, Data Processing

In Phase 5, the major processing applied to RST data for use in BT-SPAS-X includes:

- Filtering species to Chinook
- Removing fish identified as adults
- Removing yearlings
- Removing adipose clipped fish (except on Butte Creek as these were identified to not be hatchery fish)
- Removing incomplete years

Detailed information on the process of selecting years to include is available on process of selecting years to include is available on SRJPEdata on the Articles tab. Length-at-date rulesets were developed through workshops with Stream Teams to distinguish young of the year and yearlings in the RST catch data by plotting fork length and catch date for all years of available data (Vizek et al. 2025). Detailed information on the process of setting rulesets to define lifestages is available at SRJPEdata on the Articles tab. BTSPAS-X uses a weekly timestep, so data were aggregated by Julian week.

Covariate development for spring-run JPE modeling is a phased process that began with identifying meaningful covariates that could be generated from available datasets for use in models initially designed to predict historical abundances (Vizek et al. 2025). Detailed information on the process of generating covariates for stock recruit modeling process of generating covariates for stock recruit modeling is available at SRJPEdata Articles tab.

As the ultimate goal of the spring-run JPE is to develop forecast models (i.e., models that predict outmigrant abundance prior to or during the annual outmigration), we used the information learned through initial stock recruit model exploration to develop covariates that could be used in forecast versions of models (Vizek et al. 2025). Detailed information on the process of generating covariates for forecast modeling is available at SRJPEdata on the Articles tab.

3 Updating and Maintaining Datasets

3.1 Data Tackle

Monitoring data for spring-run JPE modeling continues to be collected, and the data infrastructure for the spring-run JPE was designed to continuously update datasets with these new data as seamlessly as possible. The heart of the spring-run JPE Data Management Program is DataTackle, a tablet-based digital data collection application designed by FlowWest in collaboration with the Stream Teams to ensure rapid reporting and ongoing instantaneous compatibility of data across all programs that employ the application.

DataTackle was originally designed for RST data but is being updated and expanded for collecting and reporting other data types. Data are collected in the field using DataTackle and then posted to a cloud-based database in near real-time. These data are then pulled into the SRJPEdata R package for use in models. The proposed protocol involves updating data on EDI weekly and to SRJPEdata biweekly for use in the in-season outmigrant models. Monitoring programs on Deer and Mill creeks, and on the Feather and Yuba rivers currently use DataTackle to collect RST data, and programs on Clear and Battle creeks and on the upper Sacramento River at Red Bluff Diversion Dam are working with FlowWest on DataTackle updates consistent with their program needs. Plans are underway to push DataTackle updates to participating programs via the Apple App Store. Streams that have not yet transitioned completely to DataTackle use an EDI update process on a biweekly schedule. Adult data will be updated on EDI and SRJPEdata on an annual basis.

During the first two years of producing spring-run JPE estimates, data will be checked on a quarterly basis with Stream Teams to ensure data integrity and to quickly identify issues as they may arise. As the process becomes more streamlined there will no longer be a need for regular data checks at this frequency.

3.2 Dataset Update Process to Support Models

Spring-run JPE models are envisioned to be adaptive and incorporate the best available science. As new data are collected and our understanding of the system continues to evolve, spring-run JPE models will be updated to reflect these changes. Anticipated updates to spring-run JPE models range from simply refitting coefficients for current models with updated datasets to testing new model structures and new covariates as they are developed and the Chinook salmon ecosystem and our understanding of it evolves.

Following is the proposed workflow to update datasets to support model updates and to support regular running of the models to produce spring-run JPE forecasts. More information specific to expected model update processes is described in a

Chapter 3 of these review materials. All scripts associated with these updates are version-controlled and maintained on GitHub at the SRJPEdata and SRJPEmodel repositories. In general, datasets will be updated as described below for the expected types of model updates and associated timescales.

3.2.1 Annual End of the Outmigration Season Updates (June–July)

At the end of each outmigration season, and prior to the initiation of the next juvenile outmigration season, model coefficients will be updated using updated datasets. All data sources that are pulled in and integrated within the SRJPEdata R package are updated regularly on EDI as data are logged. The main goal of the SRJPEdata package is to make data available for spring-run JPE modeling so the data would only need to be updated when they are needed for modeling. A longer-term goal would be to set up automated updates to keep the SRJPEdata package as up-to-date as possible. Currently, at the end of the monitoring season, a new year of RST catch, genetics, and survival data will be available in the SRJPEdata package by July. The following models can then be updated with the new data:

- BT-SPAS-X (Chapters 4 and 5)
- Probabilistic length-at-date (PLAD) (Chapter 6)
- Spring-run juvenile survival and travel time (Chapter 9)

3.2.2 Annual End of Adult Upstream Migration/Spawning Season Updates (December–January)

Unlike the regular RST EDI packages, adult packages are only updated annually; therefore, there is not an automated pipeline for adult data updates as there is for RST data. Data update protocols exist for each monitoring program's EDI package, and the Data Management Team will work with the Data Stewards to update the adult EDI packages as soon as possible each season, which will typically occur in December. After the EDI packages have been updated (or if an EDI package does not yet exist, data have been provided by the Stream Teams) these data will be pulled into the SRJPEdata package for modeling. At this point the stock recruit models can be updated.

3.2.3 Within Juvenile Outmigration Season Updates (January–April)

Some models require datasets to be updated during the outmigration season. Updates to RST, genetic, and some physical covariate datasets will be repeated as often as weekly during the outmigration season for the PLAD and BT-SPAS-X models. These datasets and perhaps additional physical covariate datasets will be updated on February 1, March 1, and April 1 for the in-season outmigrant and

integrated JPE models, which are the current dates configured for spring-run JPE forecasts. These same steps will be taken for new spring-run JPE forecast dates as they are established to support specific minimization measures.

4 Web Links

4.1 SRJPE on GitHub

Articles on References tab
https://srjpe.github.io/SRJPEdata/reference/index.html

Environmental data on References tab
(https://srjpe.github.io/SRJPEdata/reference/environmental_data.html

Generating covariates for stock recruit modeling
https://srjpe.github.io/SRJPEdata/articles/sr_covariates.html

Process of generating covariates for forecast modeling
https://srjpe.github.io/SRJPEdata/articles/forecast_covariates.html

Process of selecting years to include
https://srjpe.github.io/SRJPEdata/articles/years_to_include_analysis.html

Setting rulesets to define lifestages
https://srjpe.github.io/SRJPEdata/articles/lifestage_ruleset.html

Summary of data used for modeling
https://srjpe.github.io/SRJPEdata/articles/summary_of_data_for_modeling.html

SRJPE system-design-docs
https://github.com/SRJPE/system-design-docs/blob/main/srjpe_data_management_plan.md

SRJPEdata site overview article
https://srjpe.github.io/SRJPEdata/articles/site_overview.html

SRJPEdata subdomain
https://srjpe.github.io/SRJPEdata/

SRJPEmodel subdomain
https://github.com/SRJPE/SRJPEmodel

5 References

Allison, A, M Johnson, J Nichols, C Campos, J Kindopp, S Holley, V Kollmar, N Gephart, and B Harvey. 2021. *Interim Monitoring Plan for the Spring-Run Chinook Salmon Juvenile Production Estimate Science Program*. Prepared by California Department of Fish and Wildlife. Prepared for California Department of Water Resources. September. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/State-Water-Project/Endangered-Species-Protection/SR-JPE-Interim-Monitoring-Plan-Final.pdf

California Department of Water Resources (DWR), California Department of Fish and Wildlife, NOAA Fisheries, US Bureau of Reclamation, Metropolitan Water District, State Water Contractors, AECOM, and ICF. 2020. *Incidental Take Permit Spring-Run Chinook Salmon Juvenile Production Estimate Science Plan 2020–2024*. December. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/State-Water-Project/Files/ITP/ITP-Spring-run-Chinook-Salmon-JPE-Science-plan-final-approved_Final_PDF_04-05-22.pdf.

Harvey, B, P Nelson, S Gill, A Vizek, and E Cain. 2022. *Data Management Strategy for the Spring-Run Chinook Salmon Juvenile Production Estimate*. October. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/State-Water-Project/Endangered-Species-Protection/ITP-SR-JPE-Data-Mgmt-Strategy-2022-09-07.pdf.

Tables and Figures

DRAFT | Peer Review Purposes Only | Not for Citation

Tables

Table 1. Stream with Agency Affiliations and Stream Team Members

Table 1 lists Stream Team members by stream with agency affiliation in parentheses. Stream Teams collect and steward data used in spring-run Chinook salmon (*Oncorhynchus tshawytscha*) (spring-run) juvenile production estimate (JPE) models.

Stream	**Stream Team Leads & Data Stewards**
Battle Creek (US Fish and Wildlife Service)	Natasha Wingerter, Gabriella Moreno, Sam Provins, Teresa Urrutia
Butte Creek (California Department of Fish and Wildlife)	Anna Allison, Grant Henley
Clear Creek (US Fish and Wildlife Service)	Natasha Wingerter, Gabriella Moreno, Sam Provins, Teresa Urrutia
Deer Creek (California Department of Fish and Wildlife)	Ryan Revnak
Feather River (DWR)	Jason Kindopp, Kassie Henley
Lower Feather River (California Department of Fish and Wildlife)	Anna Allison, Gabriel Loera
Mill Creek (California Department of Fish and Wildlife)	Ryan Revnak
Yuba River (DWR)	Casey Campos
Sacramento River (California Department of Fish and Wildlife, US Fish and Wildlife Service)	Anna Allison, Nick Bauer, Drew Huneycutt, Corey Fernandez, David Custer, Bill Poytress

Table 2. Data Types and Model Use

Table 2 lists data types and use in models.

Data Type	Model Use
Rotary Screw Trap (RST)	BT-SPAS-X; BT-SPAS-X output is used for probabilistic length-at-date (PLAD), and both BT-SPAS-X and PLAD output based on RST catch are used to produce response and/or predictor variables for in-season outmigrant and stock recruit models.
Adult Spawner (passage, holding, carcass, redd)	Stock recruit, and potentially future iterations of PLAD.
Acoustic Tag	Juvenile survival and travel time.
Genetics	PLAD; PLAD output based on genetics is used to produce response and/or predictor variables for in-season outmigrant and stock recruit models.
Physical Conditions (flow, temperature)	BTSPAS-X, PLAD, stock recruit, juvenile survival and travel time, in-season outmigrant (all models).

Table 3. Individual Datasets by Location and Data Type

Table 3 describes individual datasets by location and data type used in spring-run JPE modeling.

Data Type	Location	Time Period	Data Access Link	Submodel Used In	Caveats & Considerations
RST	Battle Creek	2004–ongoing	Environmental Data Initiative (EDI)	BT-SPAS-X	There are two sites on Battle Creek. The upper site is used in modeling because the barrier weir separates fall- and spring-run spawning
RST	Butte Creek	1996–ongoing	EDI	BT-SPAS-X	The off-channel diversion or fyke trap operates concurrently with the RST; however, data for sampling days in which only the fyke trap was operating are not included.
RST	Clear Creek	2003–ongoing	EDI	BT-SPAS-X	-There is an upper and lower Clear Creek site. Both sites were explored for use and the upper site was eventually selected for the integrated modeling.
RST	Deer Creek	1995–2010; 2023–ongoing	Historical, Current	BT-SPAS-X	-
RST	Feather River	1998–ongoing	EDI	BT-SPAS-X	There are multiple sites on the Feather River though by design one site in the low-flow channel and one site in the high-flow channel are operated simultaneously. The low-flow channel sites were grouped for some analyses to increase sample size and the high-flow channel sites were similarly grouped.
RST	Lower Feather River	2022–ongoing	EDI	BT-SPAS-X (future)	The Lower Feather River site does not currently have enough years of data collection to be included in modeling. The plan is to include these data in the future.
RST	Mill Creek	1996-2010; 2023–ongoing	Historical, Current	BT-SPAS-X	-
RST	Yuba River	2000–2009; 2023–ongoing	EDI	BT-SPAS-X	-
RST	Lower Sacramento River (Delta Entry)	1994–ongoing	Knights Landing, Tisdale, Delta Entry	BT-SPAS-X	The Delta Entry site does not currently have enough years of data collection to be included in modeling. The plan is to include these data in the future.
Redd; Upstream Passage Estimates	Battle Creek	Redd: 2001–ongoing	Data publication in progress	Stock recruit; P2S	-
Carcass (Post-Spawn) Estimates	Butte Creek	2001–ongoing	Available on GrandTab	Stock recruit	Butte Creek also conducts holding surveys and upstream passage monitoring. Carcass estimates, developed through mark and recapture data collection, was identified as the most representative data type.
Redd; Upstream Passage Estimates	Clear Creek	Redd: 2002–ongoing	Data publication in progress	Stock recruit; P2S	-
Holding Surveys; Upstream Passage Estimates	Deer Creek	Holding: 1992–ongoing Upstream: 2014–ongoing	EDI	Stock recruit; P2S	-
Broodstock	Feather River	2014–ongoing	In progress	Stock recruit	Differentiating between spring and fall run adults in Feather River has historically been challenging. To address this we use the number of fish tagged for broodstock at the hatchery in the spring minus the number that return in the fall to represent in-river spring-run. Any over summer mortality is also removed to represent the number of spawners.
Redd; Upstream Passage Estimates	Mill Creek	Redd: 2009–ongoing Upstream: 2007–ongoing	EDI	Stock recruit; P2S	-
Upstream Passage Estimates	Yuba River	2004–ongoing	EDI	Stock recruit	Please see the Methods document available on EDI for more details on run differentiation methods.
Acoustic Tagging Data	Mill Creek, Deer Creek, Butte Creek, and Feather River	2018–ongoing	CalFishTrack Central Valley Enhanced Acoustic Tagging Project	Juvenile survival and travel time	-

Figures

Figure 1. Map of Rotary Screw Trap Monitoring Sites

Figure 1 is a map of RST monitoring sites currently used in spring-run JPE modeling. Note that the Parrott-Phelan Diversion Dam monitoring location has one RST and one fyke trap.

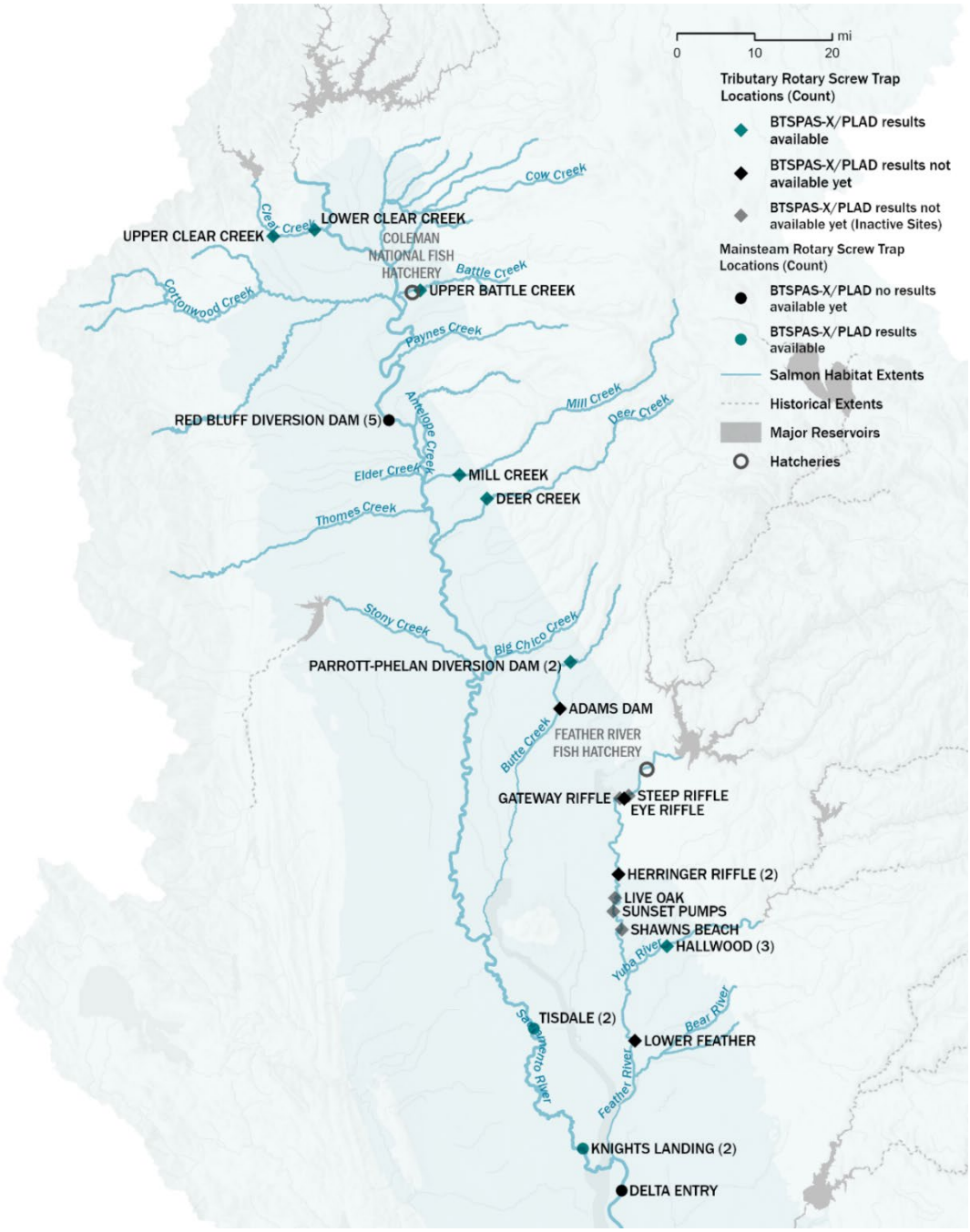

Figure 2. Map of Adult Survey Locations and Video Monitoring Stations

Figure 2 is a map of adult survey locations and video monitoring stations. Note that different types of adult data are collected on each stream and the survey reaches are labeled with the type of data collected.

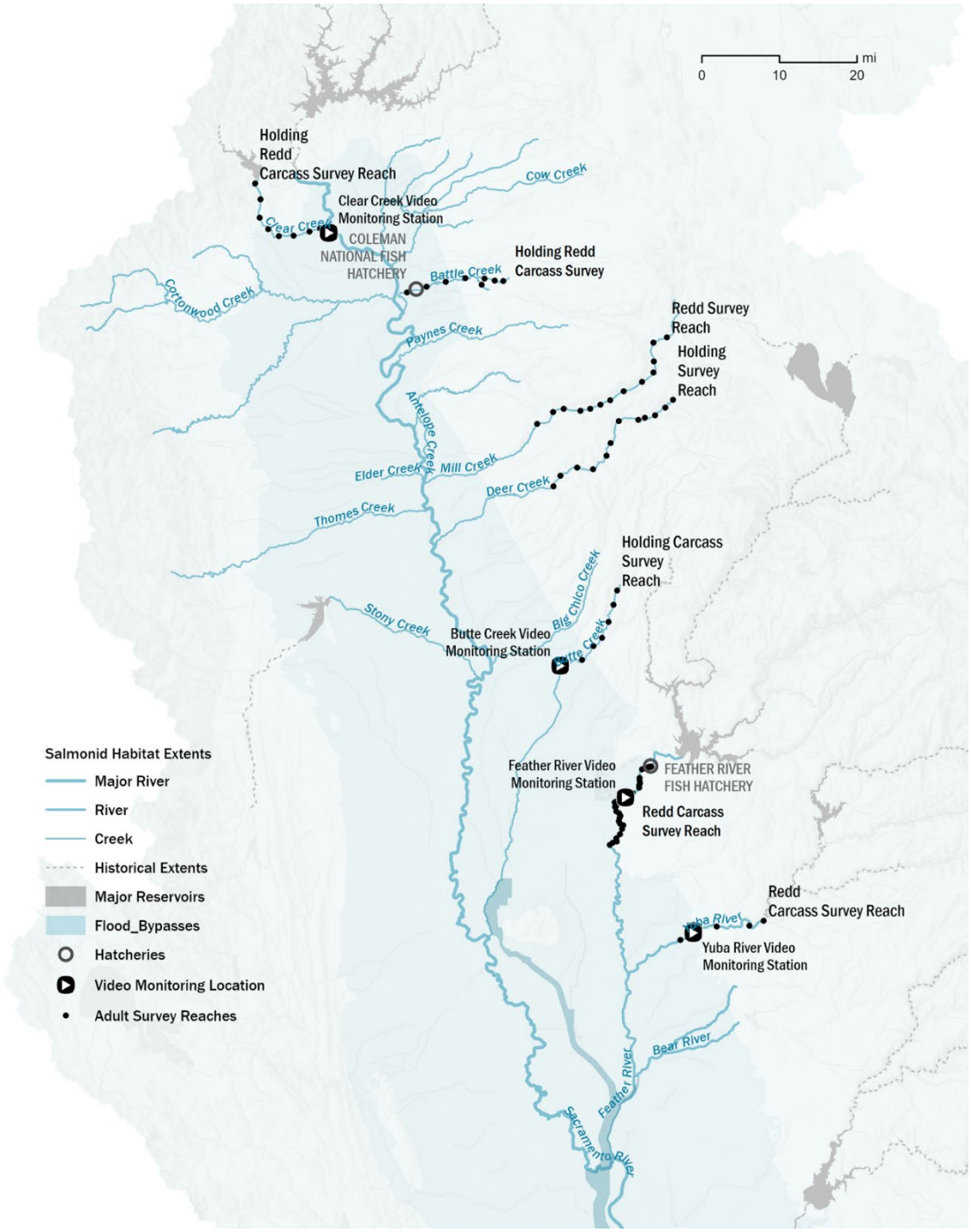

Figure 3. Map of Flow and Temperature Gage Locations

Figure 3 is a map of flow and temperature gage locations used in spring-run JPE modeling. Note that flow data for the Feather River high-flow channel is not represented on the map and is an additive value using data from gages at the hatchery, Fish Barrier Dam, and Thermalito Afterbay.