Yuba-Feather Forecast Informed Reservoir Operations

Final Viability Assessment

Appendices

February 2025

Table of Contents

Appendix A – Water Resources Engineering

A.1 Hydrologic Engineering Management Plan (HEMP)

Efforts to improve the coordinated operations of Oroville (ORO) and New Bullards Bar (NBB) dams formally began in 2006 with the Forecast-Coordinated Operations (F-CO) Program. That program has been tremendously successful in developing a common operating picture for reservoir operators, improving the observation network, and integrating single-value and, more recently, ensemble streamflow forecasts into the coordinated decisions process.

The Forecast-Informed Reservoir Operations (FIRO) program for the Yuba-Feather system is an extension of the F-CO effort and leverages the experience of FIRO efforts for Lake Mendocino and Prado Dam. The FIRO effort introduces research to improve forecasts and formally integrates streamflow forecasts into the water management decision process (Water Control Plan for the US Army Corps of Engineers (USACE) or Section 7 dams). An inter-agency inter-disciplinary steering committee (SC) was formed for the Yuba-Feather FIRO Project in June 2019.

The objective of this Hydrologic Engineering Management Plan (HEMP) is to identify through appropriate detailed technical analyses and other considerations candidate FIRO strategies for ORO and NBB dams, along with how they might be implemented in real-time operation by USACE, State Water Project (SWP) and Yuba Water Agency (YWA). A second HEMP will be developed to develop and manage system operations that meet the objectives of the F-CO Program.

The California Department of Water Resources (DWR) State Water Project (SWP) completed a Comprehensive Needs Assessment (CNA) for Oroville Dam resulting from the 2017 Oroville Dam spillway incident. Information and recommendations from this assessment have been integrated into this document.

YWA is in the process of adding a new water control structure to NBB Dam that will dramatically improve the capacity to release stored water more quickly and at lower storage levels. This analysis assumes the conditions associated with this completed construction project.

This HEMP is managed by the Yuba-Feather FIRO SC. To be consistent with USACE guidance for conduct of similar technical studies the SC prepared this HEMP as *…a technical outline of the hydrologic engineering studies necessary to formulate a solution to a water resources problem* (*Engineering Pamphlet 1110-2-9*).

This HEMP includes the following:

1. Statement of objective and overview of technical study process to provide information needed for this assessment.
2. Identification of tasks to be completed for the technical analysis. (Table A-1).
3. Identification of candidate FIRO alternatives to be analyzed. (Table A-2).
4. Specification of requirements for all FIRO alternatives that will be considered. (Table A-3).

5. Identification of hard criteria as well as project and system-wide considerations. (Tables A-4, A-5, A-6).
6. Identification of initial tentative performance metrics for FIRO alternative evaluation. (Table A-7).
7. Identification of the project team members and their roles and responsibilities for conducting, reviewing, and approving of the hydrologic engineering study. (Tables A-8, A-9, A-10).
8. Risks to the success of this study and mitigation actions are shown in Table A-11.

Objective of Technical Analysis, Overview of Process, and Tasks to be Completed

The objective of the hydrologic engineering study described herein is to identify and evaluate FIRO alternatives for ORO and NBB dams in a systematic, defendable, repeatable manner, thus providing information to the SC so that it may identify the best FIRO strategy for NBB Dam.

The process used to meet the hydrologic engineering study objective is a "nominate-simulate-evaluate-iterate" process, consistent with the process used commonly by USACE for water resources planning studies. Tasks in this process, as applied for technical analyses to support the ORO Dam CAN and the NBB Dam FIRO Viability Assessment, and include the following:

1. A set of feasibility criteria and performance metrics is developed for assessing and comparing FIRO alternatives. This set will be applied to all alternatives, thereby permitting the project delivery team (PDT) to compare and rank alternatives for consideration by the SC.
2. A set of alternative FIRO strategies is nominated by the PDT. The strategies are screened to ensure they meet specified requirements, which are described below.
3. Performance of the river-reservoir system with each FIRO strategy is simulated using a common set of meteorological and hydrological conditions. HEC-ResSim more likely will act as the "gatekeeper" for all alternatives to ensure that the physical constraints and attributes of the system are consistently applied.
4. Simulation results are used to evaluate the viability and performance of each strategy. The evaluation uses metrics identified in Task 1, comparing each alternative to performance for the *without-project* (baseline) condition, which is operation following the water control plan (WCP) included in the current water control manual (WCM). If results of the evaluation inform refinements to FIRO strategies, the simulation and evaluation tasks are repeated with enhanced strategies to the extent that resources allow.
5. The PDT uses the technical analysis results to rank the alternatives and submits the rankings to the SC for consideration.

These tasks are described in more detail in Table A-1. Major tasks are listed in column 1 and subtasks in column 3.

Major Task (1)	**Description (2)**	**Subtasks (3)**

Task 1. Select performance metrics	Both quantitative and qualitative measures of performance will be identified. Methods of computation of quantitative measures will be described.	**Task 1.1.** With appropriate input from subject matter experts, formulate candidate set of quantitative and qualitative measures of performance. Define methods for assessing these for typical FIRO strategies. Screen set to select feasible metrics for ALL likely alternatives to permit objective comparison of strategies. Prepare technical memo. Submit to SC for review. **Task 1.2.** Receive comments from SC. Revise selected set of performance metrics as required. **Task 1.3**. If necessary, design, develop, and test software applications (scripts, spreadsheets, etc.) to apply selected metrics.
Task 2. Nominate/formulate alternative FIRO strategies that will be considered	Each alternative FIRO strategy to be considered will be identified and described, along with the method by which performance with the strategy will be evaluated.	**Task 2.1**. With appropriate input from subject matter experts, formulate candidate set of FIRO strategies to be considered. Describe each strategy in memo, submit proposed list/memo to SC for approval. **Task 2.2.** Receive comments from SC and revise list as appropriate. Get SC agreement to proceed with comparison. **Task 2.3**. Identify software applications that will be used to model FIRO strategies.
Task 3. Side studies	Identify, conduct, document, and incorporate outcomes of "side studies" that affect the simulation and evaluation of alternatives.	**Task 3.1.** Identify any additional "side studies" that must be completed to provide information required for simulation. Details of side studies will be identified in this subtask, with scope of work and schedule submitted to SC for approval. **Task 3.2.** Undertake and complete side studies, as approved by SC. Document findings. Incorporate findings in selected FIRO strategy models or procedures.
Task 4. Simulate performance with each alternative	Each alternative FIRO strategy will be simulated with the HEC-ResSim model with a consistent set of hydrologic boundary conditions and system constraints (identified in Table 3).	**Task 4.1.** Considering all FIRO strategies to be evaluated, identify boundary conditions and initial states of the system to be considered in simulation for comparison. Document. **Task 4.2.** Simulate performance of ORO and NBB dams with candidate strategies. Prepare technical memo describing application of each strategy. Prepare database of results (for use in Task 5).
Task 5. Using results of simulation, evaluate each alternative in terms of identified	Each alternative FIRO strategy will be analyzed and the appropriate performance metric statistics computed.	**Task 5.1.** Using database of results from the HEC-ResSim simulation of each FIRO strategy (from Task 4.2) apply software applications (scripts, spreadsheets, etc.) from Task 1.3 to compute performance metrics for each strategy. **Task 5.2.** Revise FIRO strategies and performance metrics as necessary to ensure fair,

performance metrics		repeatable comparisons. This subtask acknowledges initial uncertainty about compatibility of strategies and metrics. **Task 5.3.** Document results of evaluation in technical memo.
Task 6. Compare the alternatives by comparing the metrics	Each alternative FIRO strategy evaluation will be compared against the baseline and against each other.	**Task 6.1.** Using results from Task 5, prepare charts, tables, etc. to compare performance of strategies. Prepare technical memo with this information and submit to SC for information. **Task 6.2.** Refine strategies if evaluation and comparison expose opportunities for "quick gains" through minor adjustments to strategies. Repeat subtasks Task 4.2 through Task 5.1 with revised results. **Task 6.3.** Prepare final technical memo on simulation, evaluation, and comparison. Submit for SC review. Receive SC comments and revise technical memo as needed.
Task 7. Brief SC on findings and facilitate the selection of a preferred approaches to be refined in the FVA	Each alternative FIRO strategy comparison will be scrutinized, a preferred approaches and refinements for the FVA identified and documented and presented to the SC.	**Task 7.1.** Using results of comparison from Task 6, rank alternatives considering individual metrics from Task 1. Document findings. **Task 7.2.** Provide comparisons and ranking to SC. **Task 7.3.** Document recommended refinements for the FVA process.

Table A-1. *Tasks and Subtasks to be Completed for Hydrologic Engineering Study of FIRO Strategies*

FIRO Alternatives to be Evaluated

Selection of candidate FIRO alternatives has been completed by the Water Resources Engineering (WRE) Team (Task 2). These candidate alternatives were delivered to the Corps WCM Update Team and will be evaluated through the procedures defined in this document. The existing WCM operations for both ORO and NBB will also be evaluated to establish the performance baseline. Table A-2 shows the list of WCP alternatives to be evaluated.

ID	Dam	Alt	Alt Description	Operation Principle
1	ORO	EO	Existing Operations	Exiting WCP from current WCM.
2	ORO	PresFcst_1	Use best-estimate forecast volumes to inform guide curve TOC computation and inflow-based releases.	Relies on an elevation-based guide curve that is computed based on forecast inflow volumes. When in the flood control pool, intent is to evacuate the storage in a controlled manner to reduce downstream peak

				flows. Stepped releases are proposed.
3	ORO	IterFcst_1	Use ensemble streamflow forecast members to determine a release based on an iterative process to maintain the same dam risk profile as current operations.	Identify a "minimally changed release" through the flood event. This release (or pattern) is identified as the maximum release that is needed to balance the use of the flood pool but not result in adverse dam safety concerns. The operation seeks to answer the question, what is the release needed to make it through this event safely? Use the forecast information, and the associated uncertainty to identify the release.
4	ORO	EFO	Ensemble Forecast Operations (EFO) Model using the full range of reservoir storage.	Manages risk of exceeding a defined critical storage threshold using a developed risk curve and ensemble streamflow forecasts. Full range of storage is available for release decisions.
5	ORO	EFO Hybrid	Hybrid EFO Model limited to a defined FIRO Space.	Manages the risk of exceeding a defined critical storage threshold using a developed risk curve and ensemble streamflow forecasts. FIRO release decisions limited to the define FIRO Space.
6	NBB	EO	Existing Operations	Existing WCP from current WCM.
7	NBB	FIRO GC	FIRO Guide Curve. FIRO for flood control and water supply using a forecast-based guide curve to specify drawdown in advance of flood events and conditional storage of water in the gross pool when forecast is dry.	Evacuate volume above FIRO guide curve over less than one day time window. Increases storage utilization to mitigate high downstream flood releases.
8	NBB	FIRO RS	FIRO Release Schedule. FIRO for flood control using a forecast-based release schedule to specify drawdown in advance of flood events.	Evacuate conservation space to absorb forecast event, reducing peak releases and peak storage in NBB.

9	NBB	EFO	Ensemble Forecast Operations (EFO) Model using the full range of reservoir storage.	Manages risk of exceeding a defined critical storage threshold using a developed risk curve and ensemble streamflow forecasts. Full range of storage is available for release decisions.
10	NBB	EFO Hybrid	Hybrid EFO Model limited to a defined FIRO Space.	Manages the risk of exceeding a defined critical storage threshold using a developed risk curve and ensemble streamflow forecasts. FIRO release decisions limited to the define FIRO Space.

Table A-2. *List of WCP alternatives to be evaluated.*

Requirements of all candidate strategies are shown in Table A-3. Tables A-4, A-5, and A-6 show additional constraints and objectives that should be met by all the alternatives. The operational considerations in Tables A-5 and A-6 are used to create the evaluation metrics provided in Table A-7.

ID	Description
1	The candidate FIRO strategy must satisfy all relevant USACE engineering regulations (ERs), including, but not limited to, the following: • ER 1105-2-100 *Planning Guidance Notebook* • ER 1105-2-101 *Risk Assessment for Flood Risk Management Studies* • ER 1110-2-240 *Water Control Management* • ER 1110-2-1156 *Safety of Dams Policy and Procedures* • ER 1110-2-1941 *Drought Contingency Plans* • EM 1110-2-3600 *Management of Water Control Systems* • ER 1110-2-8156 *Engineering and Design Preparation of Water Control Manuals* • EM 1120-2-1420 *Engineering Requirements for Reservoirs*
2	The analytical tools required for implementation of the candidate FIRO strategy must be compatible with the USACE's Corps Water Management System (CWMS) software. In addition, results of any analyses completed with software not currently certified for use by USACE must be demonstrated to produce results consistent with USACE software results.
3	Streamflow forecasts used by the candidate FIRO strategy must be those provided by the California-Nevada River Forecast Center (CNRFC) of the National Weather Service. Simulated streamflow forecasts must be consistent with the skill characteristics of those issued by the CNRFC. As appropriate for the alternative, the forecast used can be ensemble and/or single value.
4	The FIRO strategy must satisfy the hard (inviolable) operation constraints shown in Table 2.
5	The FIRO strategy should represent, and to the extent possible, meet the operation objectives shown in Tables 3 and 4.

6	Software development needed to implement the FIRO alternative must be limited for the Viability Assessment, as the objective is to select from amongst a set of readily available (or nearly so) strategies.
7	Simulations should be computed at an hourly time step.

Table A-3. *Requirements of all alternative WCP strategies*

ID	Limiting Condition	Description
1	Satisfy ORO Water Control Manual Flood Control Diagram	Meet all specific requirements stated on current Flood Control Diagram
2	Satisfy ORO Water Control Manual Emergency Spillway Release Diagram (ESRD)	Meet all specific requirements stated on current Emergency Spillway Release Diagram (ESRD)
3	Satisfy NBB Water Control Manual Flood Control Diagram	Meet all specific requirements stated on current Flood Control Diagram
4	Satisfy NBB Water Control Manual Emergency Spillway Release Diagram (ESRD)	Meet all specific requirements stated on current Emergency Spillway Release Diagram (ESRD)
5	Do not assume Marysville Dam is in place	The 1972 WCM operation assumes storage is available in Marysville Reservoir. Marysville Reservoir was never built.
6	Satisfy release rate of change constraints associated with increases and decreases	As documented
7	Include function of new NBB secondary spillway	The FIRO alternatives must incorporate the function of the new NBB secondary spillway

8	Do not require other than currently available streamflow forecasts	CNRFC deterministic and ensemble streamflow forecasts are available up to 4 times per day during major runoff events. For evaluation purposes, forecast updates will be once per day.

Table A-4. *Hard (Inviolable) Operational Constraints that Must be Satisfied by All FIRO Strategies*

ID	Operational Consideration	Description
1	Reduce the frequency of critical release exceedance from ORO and NBB	Alternative should decrease the frequency of critical releases from both dams
2	Reduce the frequency of ORO releases that result in more than 180,000 cfs in the Feather River at Yuba City	Maximum F-CO flow target for ORO
3	Reduce the frequency of NBB releases that result in more than 180,000 cfs in the Yuba River at Marysville	Maximum F-CO flow target for NBB
4	Reduce the frequency of releases from ORO and NBB that result in more than 300,000 cfs in the Feather below Yuba City and 320,000 cfs in the Feather River below the Bear River.	Combined F-CO flow targets for ORO and NBB

5	Avoid negative impacts to spring refill	Alternatives should not reduce the ability of ORO and NBB to meet water supply delivery objectives
6	Avoid the use of the ORO emergency spillway	Operational objective for dam safety
7	Avoid negative impacts on hydropower generation	Hydropower production should be maintained or possibly enhanced
8	End of flood season storage	Consider the effect of FIRO operation on storage at the end of the flood season (through the end of May).

Table A-5. *Operational Considerations that Should be Evaluated in the Hydrologic Engineering Study.*

ID	Operational Consideration	Description
1	Implementation of F-CO of Lake Oroville and NBB Reservoir	Consider and support the existing YF F-CO program.
2	Operational resiliency	The FIRO alternative should be resilient to a wide range of hydrologic events within the watershed. For example, the operation should be resilient to a range of storm-centering and events of key frequencies occurring within the Yuba and Feather watersheds.

Table A-6. *System-Wide Operational Considerations that Should be Evaluated in the Hydrologic Engineering Study.*

Metrics for Evaluating Viability and Efficiency of Alternatives

The efficiency of FIRO will be evaluated with a set of measurable statistics (Task 1). These will be used in the same manner (to the maximum extent possible) to assess each alternative objectively. An initial list of metrics and the manner of computing or calculating each is shown in Table A-7.

ID	Metric Description	Category	Likely Method of Computation
M1	Flood Season maximum discharge frequency from ORO Dam	Flood risk management	Frequency curve. See Simulation Plan.

M2	Flood Season maximum pool elevation frequency function of ORO Dam	Flood risk management	Frequency curve. See Simulation Plan.
M3	Flood Season maximum discharge frequency from NBB Dam	Flood risk management	Frequency curve. See Simulation Plan.
M4	Flood Season maximum pool elevation frequency function of NBB Dam	Flood risk management	Frequency curve. See Simulation Plan.
M5	Flood Season maximum flow-frequency curves at key downstream locations	Flood risk management	Frequency curve. See Simulation Plan. CVHS frequency analysis. Key downstream locations are Yuba River at Marysville, Feather River at Yuba City, Yuba and Feather River Confluence, and Feather River near Nicolaus.
M6	ORO Reservoir storage at the end of Flood Season (spring refill)	Water supply	Reservoir routing. See Simulation Plan. Include detailed metrics on potentially the following: Changes in reservoir storage levels
M7	NBB Reservoir storage at the end of Flood Season (spring refill)	Water supply	Reservoir routing. See Simulation Plan. Include detailed metrics on potentially the following: Changes in reservoir storage levels
M8	ORO Hydropower production	Hydropower management	See Simulation Plan. Changes in monthly and annual megawatt production output frequency curve.
M9	NBB Hydropower production	Hydropower management	See Simulation Plan. Changes in monthly and annual megawatt production output frequency curve.

Table A-7. *List of Metrics for Evaluation of WCP Alternatives (listed in Table A-2).*

Bookend Analysis

To better understand the maximum benefit of forecasts, all non-baseline alternatives will be configured and run with full foresight of future streamflow conditions for the full lead time of the forecasts utilized (perfect forecasts).

The “bookends” will be established by the baseline alternative and the results of the perfect forecast simulations for the FIRO alternatives for each dam. The current position between the two “bookends” will be established through the evaluation of each non-baseline alternative in Table A-2 using currently available forecasts.

Project Delivery Team Members and their Roles

The PDT for evaluation of FIRO alternatives includes subject matter experts who will complete the analyses described herein, report on the findings and understandings, and recommendations in memo form to the YF FIRO SC. This work effort is led by the YF FIRO PVA Water Resources Engineering Team. PDT members are identified in Table A-8.

- Yuba-Feather FIRO steering committee
- SWP technical staff and consultants (HDR)
- YWA technical staff and consultants (MBK)
- USACE Headquarters staff (HQ)
- USACE Engineering Research and Development Center (ERDC) staff
- USACE, South Pacific Division (SPD) staff
- USACE, Sacramento District (SPK) staff
- Center for Western Weather and Water Extremes, Scripps Institution of Oceanography at University of California, San Diego. Includes Robert K. Hartman Consulting Services (RKHCS) and Sonoma Water staff under contract to support FIRO efforts.

Table A-8. *New Bullards Bar Dam FIRO Alternatives Evaluation Technical Analysis PDT Members*

The PDT members have one of four roles, consistent with established project management planning, as shown in Table A-9. These roles vary by hydrologic engineering task. Table A-10 shows roles assigned to PDT members for the analysis described herein.

ID	Role	Description of Duties
R	Responsible	Responsible for completing the analyses described herein.
A	Accountable	Answerable for correct and thorough completion of task; ensures requirements are met; delegates work to those responsible.
C	Consulted	As SMEs, offer opinions through two-way communication with those responsible and accountable, about conduct of analyses.
I	Informed	Keep up to date on progress through two-way communication.

Table A-9. *Project Roles*

Major Task	Steering Committee	SWP/YWA Tech Staff and Consultants	USACE HQ	USACE ERDC	USACE SPD	USACE SPK	CW3E
Task 1. Select performance metrics	I	R	I	C	C	R	R
Task 2. Nominate/formulate alternative FIRO strategies that will be considered	C	R	I	C	C	R	R
Task 3. Side studies	C	R	I	C	C	R	R
Task 4. Simulate performance with each alternative	I	R	I	I	I	C	R
Task 5. Using results of simulation, evaluate each alternative in terms of identified performance metrics	I	R	I	I	I	C	R
Task 6. Compare the alternatives by comparing the metrics	I	R	I	I	C	C	R
Task 7. Brief SC on findings and	I	R	I	I	I	R	R

facilitate the selection of a preferred alternative							

Table A-10. *PDT Roles by Task*

Schedule for Completion of Technical Analysis

Figure A-1 shows the schedule for completion of the project tasks. All work on all tasks will be completed by December 31, 2021.

(To be developed).

Figure A-1. *Schedule for completion of hydrologic engineering study to recommend FIRO strategy for ORO and NBB dams.*

Risks to Success of Study

Risks to the success of this study and mitigation actions are shown in Table A-11.

Potential Failure Mode	Actions PDT can take to Mitigate
Simulation or evaluation software does not function as expected.	Limit analysis to use of software that is readily available and has been stress tested.
Necessary data—including hydrological, meteorological, water use, vulnerability—are not readily available.	Limit analysis to use of best-available data.
Key personnel are not available to complete tasks.	Ensure back up staff for all critical tasks.
Critical path tasks fall behind schedule due to unforeseeable distractions and disruptions.	Limit project activities to those that are necessary to satisfy objectives.
PDT disagrees about technical analysis procedures.	Defer to PDT project assignments (see above).
Nature of alternative FIRO strategy prevents evaluation with selected metrics.	Disqualify alternative from further consideration unless metrics can be adjusted and applied in uniform manner for all alternatives.

Table A-11. *Project Risks*

A.2 Systems Operations Presentation

System Operation

- System operation subteam goals:
  - Disambiguation of tributary constraints
  - Improved FRM performance through early FIRO release action

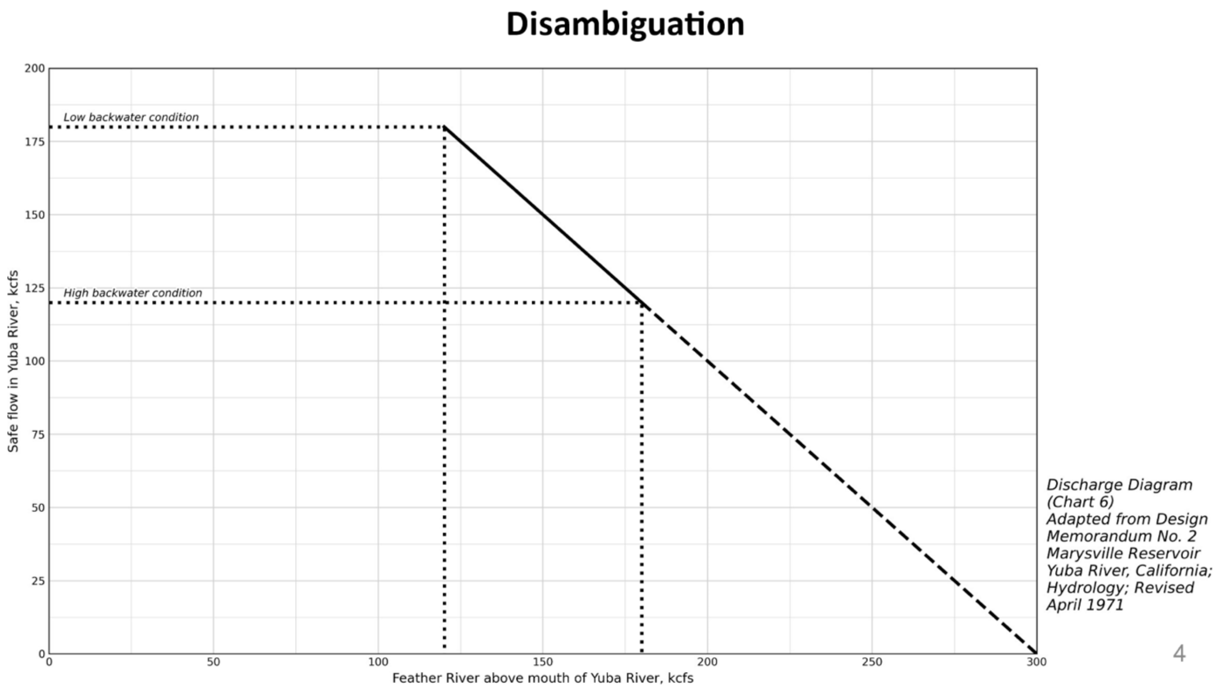
2

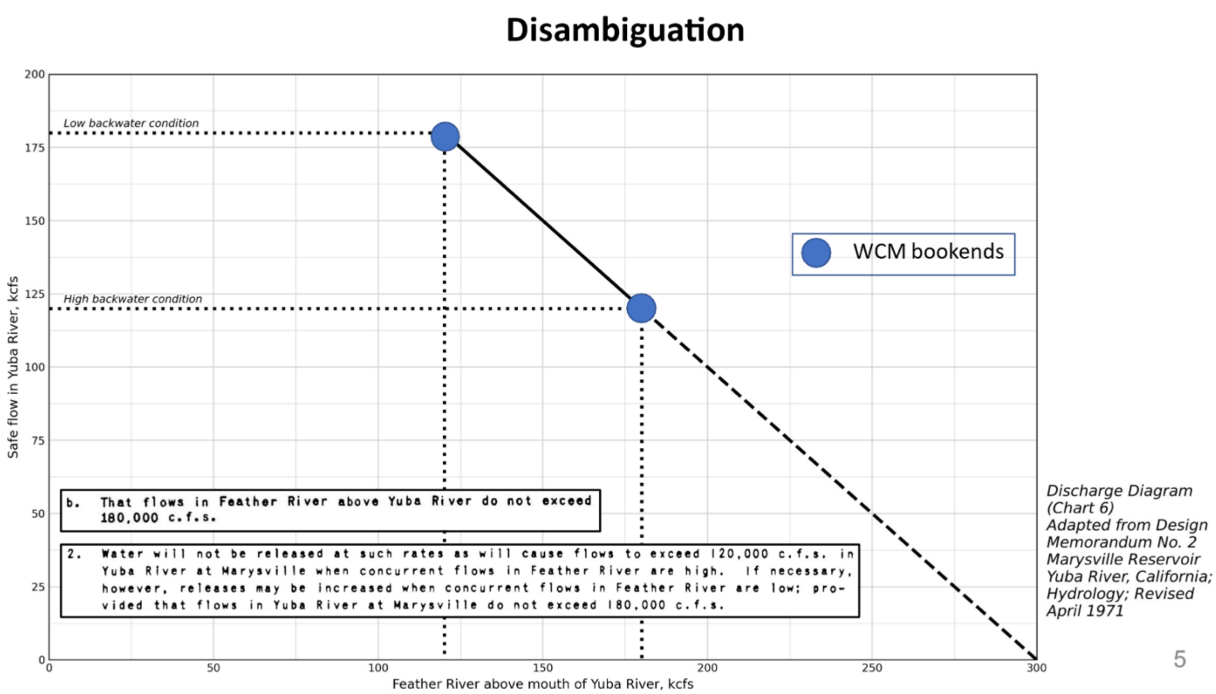
At-Site vs System Rules

PVA Alternative 1		PVA Alternative 2		PVA Alternative 3	
New Bullards Bar	**Oroville**	**New Bullards Bar**	**Oroville**	**New Bullards Bar**	**Oroville**
		FIRO GC release	**FIRO release schedule**	**EFO release**	**EFO release**
	Inflow-based release schedule		Inflow-based release schedule		
ESRD	ESRD	ESRD	ESRD	ESRD	ESRD
Rate of increase	Rate of increase	Rate of increase	Rate of increase	Rate of increase	Rate of increase
Rate of decrease	Rate of decrease	Rate of decrease	Rate of decrease	Rate of decrease	Rate of decrease
Min flow = 5 cfs	Min flow = 600 cfs	Min flow = 5 cfs	Min flow = 600 cfs	Min flow = 5 cfs	Min flow = 600 cfs
Colgate hydraulic limit	Hyatt hydraulic limit	Colgate hydraulic limit	Hyatt hydraulic limit	Colgate hydraulic limit	Hyatt hydraulic limit
Max release = 50,000 cfs	Objective release = 150,000 cfs		Objective release = 150,000 cfs		Objective release = 150,000 cfs
Max release = peak event inflow in last 120 hours					
Max flow at Marysville = 180,000 cfs	Max flow at Yuba City = 180,000 cfs	Max flow at Marysville = 180,000 cfs	Max flow at Yuba City = 180,000 cfs	Max flow at Marysville = 180,000 cfs	Max flow at Yuba City = 180,000 cfs
Max flow at Confluence = 300,000 cfs	Max flow at Confluence = 300,000 cfs	Max flow at Confluence = 300,000 cfs	Max flow at Confluence = 300,000 cfs	Max flow at Confluence = 300,000 cfs	Max flow at Confluence = 300,000 cfs
Max flow at Nicolaus = 320,000 cfs	Max flow at Nicolaus = 320,000 cfs	Max flow at Nicolaus = 320,000 cfs	Max flow at Nicolaus = 320,000 cfs	Max flow at Nicolaus = 320,000 cfs	Max flow at Nicolaus = 320,000 cfs

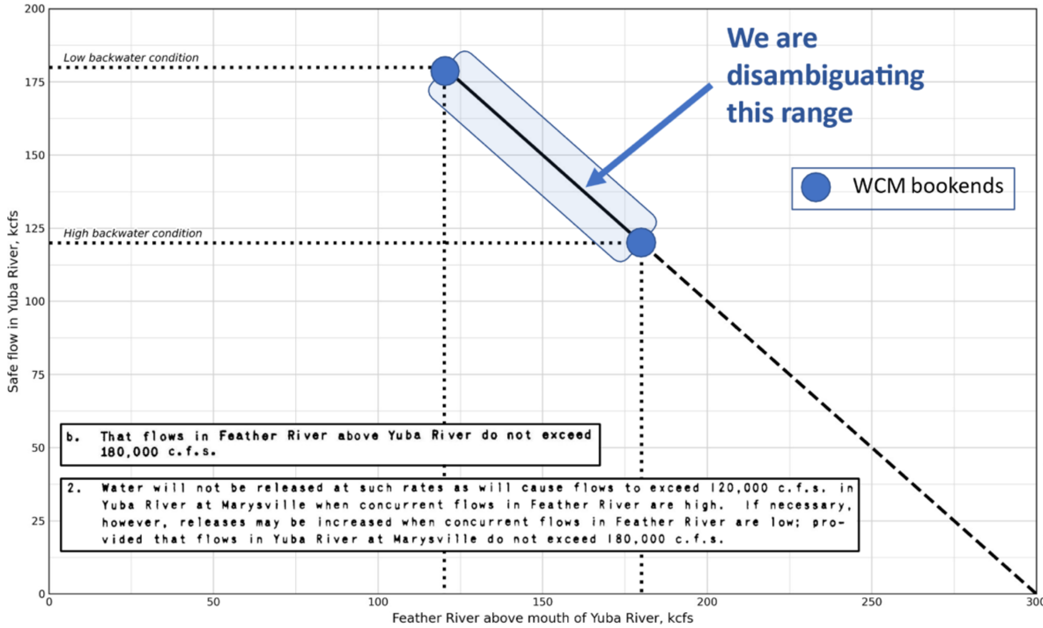
Legend
At-site operation
System operation
Unique to operations set

3


Disambiguation
Low backwater condition
High backwater condition
Safe flow in Yuba River, kcfs
Feather River above mouth of Yuba River, kcfs
0
25
50
75
100
125
150
175
200
0
50
100
150
200
250
300
Discharge Diagram (Chart 6) Adapted from Design Memorandum No. 2 Marysville Reservoir Yuba River, California; Hydrology; Revised April 1971
4


Disambiguation
Low backwater condition
High backwater condition
WCM bookends
b. That flows in Feather River above Yuba River do not exceed 180,000 c.f.s.
2. Water will not be released at such rates as will cause flows to exceed 120,000 c.f.s. in Yuba River at Marysville when concurrent flows in Feather River are high. If necessary, however, releases may be increased when concurrent flows in Feather River are low; provided that flows in Yuba River at Marysville do not exceed 180,000 c.f.s.
Safe flow in Yuba River, kcfs
Feather River above mouth of Yuba River, kcfs
0
25
50
75
100
125
150
175
200
0
50
100
150
200
250
300
Discharge Diagram (Chart 6) Adapted from Design Memorandum No. 2 Marysville Reservoir Yuba River, California; Hydrology; Revised April 1971
5

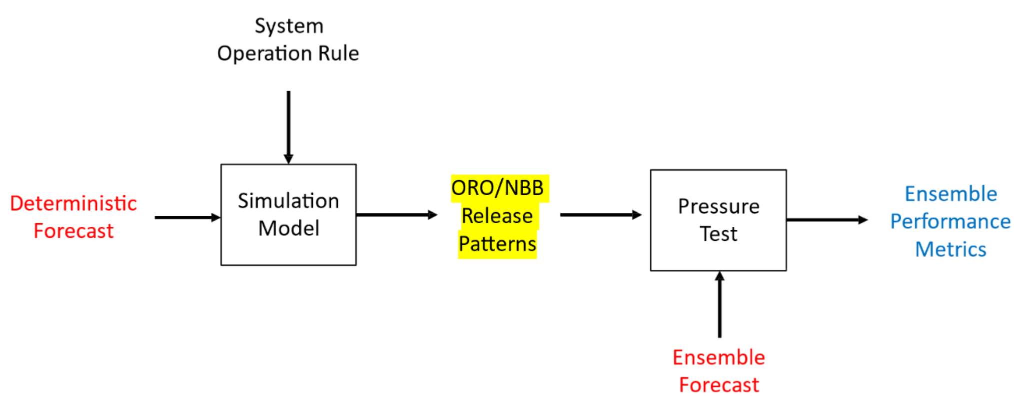
Disambiguation



Discharge Diagram (Chart 6) Adapted from Design Memorandum No. 2 Marysville Reservoir Yuba River, California; Hydrology; Revised April 1971

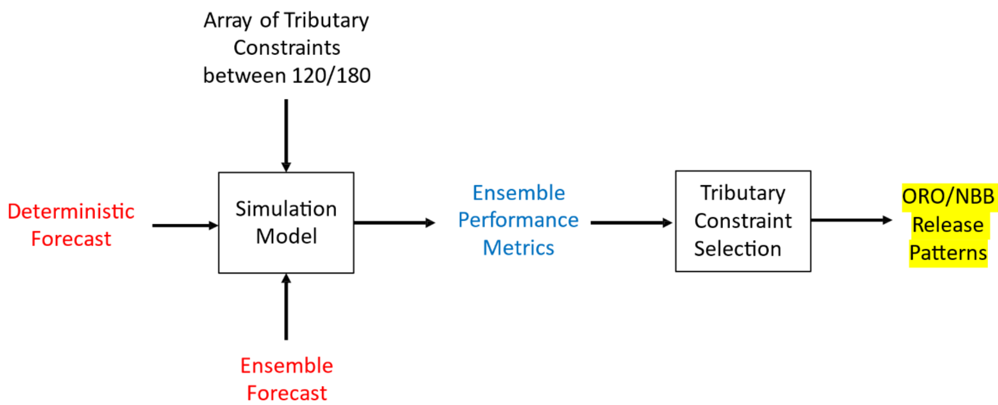
6

Current System Operation



7

Alternative System Operation



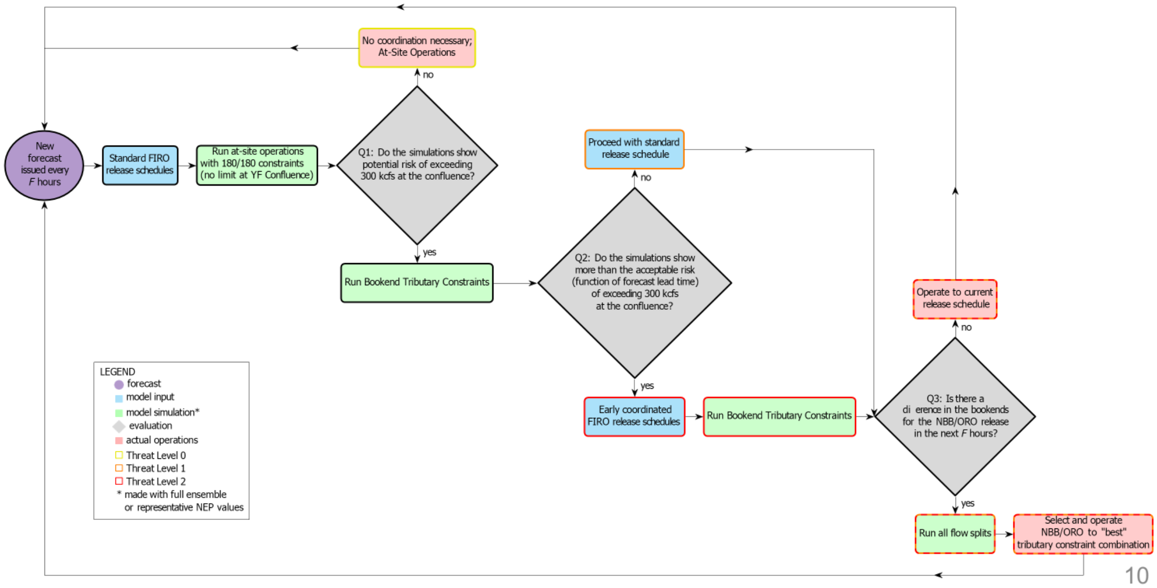
8

February 1 Workshop Summary

- Took significant steps toward a common understanding of the alternative system operation depicted in the flow chart and the simulation method used for the disambiguation of the downstream flow constraints.
- The need for further discussion on specific parts of the alternative system operation were defined and will can be included as part of the next workshop.
- Learned more from USACE about the structure of the Water Control Manual (WCM) update documentation, including the development of a **Yuba-Feather master manual** which guides the ORO and NBB WCMs and a **document** showing the **system operation (decision support system)** common to both reservoirs.
- Gathered input from agencies on key system operation **performance metrics**, including preliminary input on the **relative importance** of those metrics. This input will help in developing a candidate system operation that best achieves **risk equity**.
- Discussed performance metric balance, which revealed that the **Feather below Bear River constraint** is likely **harder to meet** than the Feather below Yuba River constraint due to its location further downstream and the additional uncontrolled flow the Bear delivers to the mainstem Feather River. The **alternative system operation** can be **structured to take this into account**.
- Agencies had **general agreement on key metrics**, thus providing encouragement that a **commonly accepted alternative system operation** is **possible**.

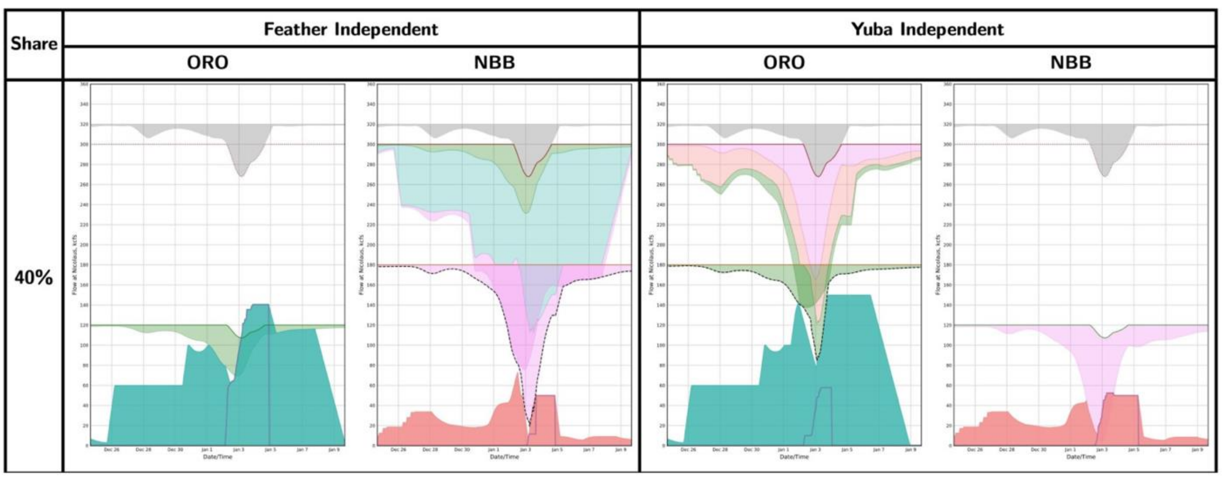
9

System Operation Alternative - Process



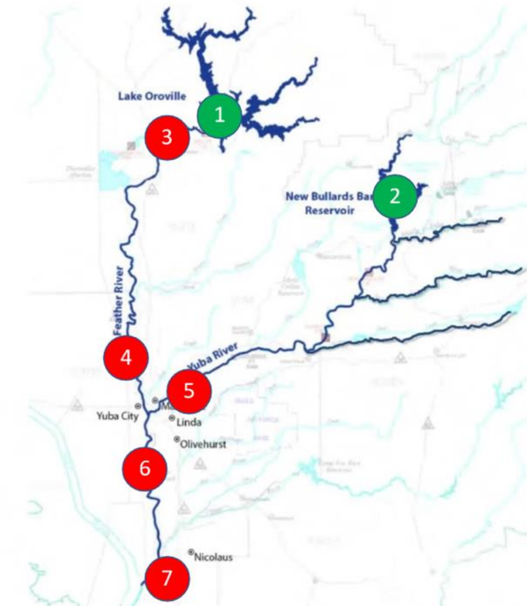
10

System Operation Alternative – Reservoir Simulation



11

Key Performance Metrics



RESERVOIR LEVELS

1. Oroville
2. New Bullards

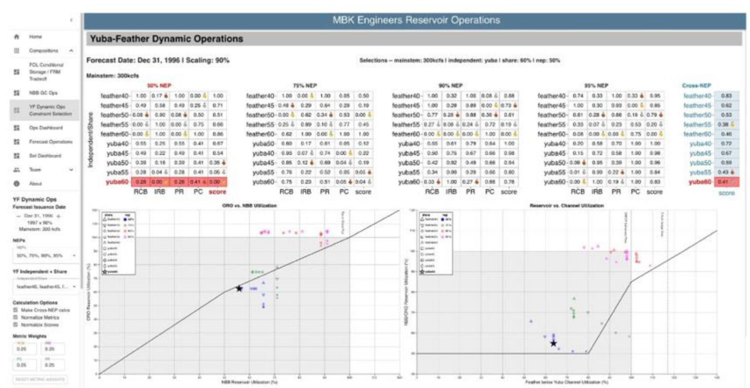
DOWNSTREAM FLOWS

3. Oroville Outflow
4. Feather @ Yuba City
5. Yuba @ Marysville
6. Feather below Yuba
7. Feather below Bear

What is "best?" – Methods

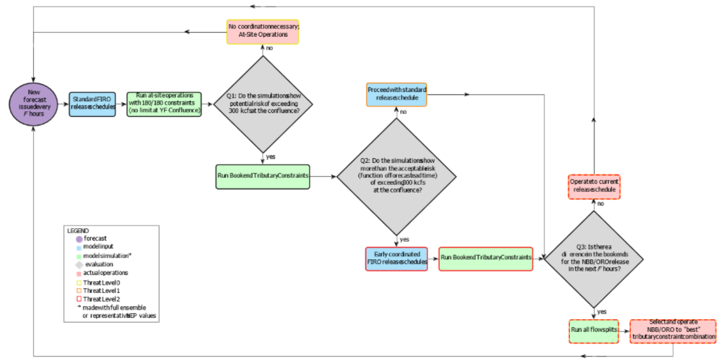
FILL IN WEIGHTS		
Metric	**Threshold**	**Weight**
Peak NBB elevation	Top of Gross Pool (1956.0 ft)	0.2
	1950.0 ft	
	1940.0 ft	
Peak ORO elevation	Emergency Spillway (901.0 ft)	0.2
	895.0 ft	
	875.0 ft	
Peak NBB outflow	Objective release (50,000 cfs)	
Peak ORO outflow	Objective release (150,000 cfs)	
Peak Feather below Yuba flow	WCM (300,000 cfs)	0.2
	Reduced (250,000 cfs)	
	Danger (66 ft/230,000 cfs)	
	Flood (65 ft/220,000 cfs)	
	Monitor (57 ft/125,000 cfs)	
Peak Feather below Bear flow	WCM (320,000 cfs)	0.2
	Reduced (270,000 cfs)	
	Danger (48 ft/260,000 cfs)	
	Flood (47 ft/204,000 cfs)	
	Monitor (39 ft/38,000 cfs)	
Peak Yuba at Marysville flow	Flow Split (180,000 cfs)	0.1
	Flow Split (120,000 cfs)	
	Danger (88 ft/103,000 cfs)	
	Flood (87 ft/95,600 cfs)	
	Hallwood evac. (83 ft/68,600 cfs)	
	Monitor (74 ft/25,900 cfs)	
Peak Feather at Yuba City flow (below Yuba for flood stages)	Danger (81.2 ft/413,700 cfs)	
	Flood (80.2 ft/378,700 cfs)	
	Flow Split (180,000 cfs)	0.1
	Flow Split (120,000 cfs)	
	Monitor (65 ft/91,000 cfs)	

Reservoir elevation weights	0.4
Downstream channel weights	0.6
WEIGHTS SUM (MUST = 1)	1

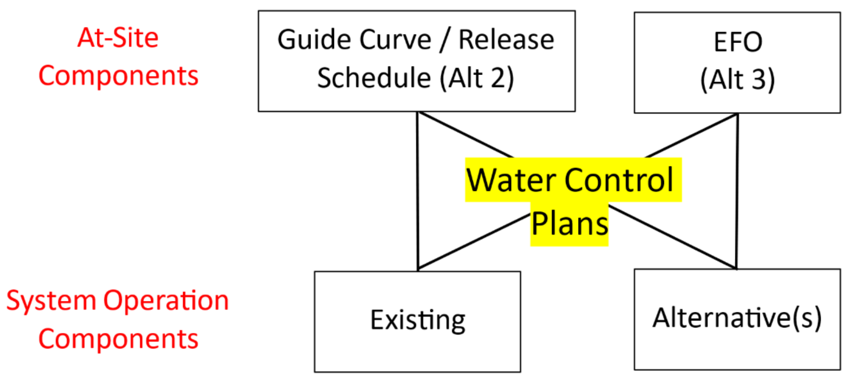


March 8 Workshop

- Alternative System Operation Process and Algorithm Updates
- Bear River Constraint Handling
- Tributary Constraint Disambiguation Methods; What is “best?”



FVA Alternatives



A.3 Simulation Plan

Memorandum

Date:	Friday, June 02, 2023
Project	Forecast-Informed Reservoir Operations (FIRO) Program
Subject:	FVA Simulation Plan

Situation

The California Department of Water Resources (DWR) and Yuba Water Agency (YWA) are participating in the Yuba-Feather Forecast-Informed Reservoir Operations (FIRO) Program, a multi-agency partnership focused on evaluating the viability of FIRO at Oroville and New Bullards Bar dams and identifying opportunities for forecast enhancement. Oroville Dam on the Feather River is owned and operated by DWR, and New Bullards Bar on the Yuba River is owned and operated by YWA. For flood control, the dams are operated separately and as a system to avoid exceeding the maximum objective flows in the Feather River below the Yuba River and in the Feather River below the Bear River. Flood operation rules for the dams are prescribed in each dam's water control manual (WCM) developed by the U.S. Army Corps of Engineers Sacramento District (SPK).

FIRO is a strategy that leverages advances in forecasting technology to allow greater flexibility in reservoir operation and in turn, to potentially enhance flood control and water supply benefits. By explicitly considering forecasted inflow in release decision making, operators can optimize release timing and magnitude to pass flood flows safely or can store water that would normally be released when no significant inflow is forecast.

Task

To assess FIRO viability, the water resources engineering (WRE) team will develop and evaluate operation alternatives that explicitly include inflow forecasts in release decision making at Oroville Dam and New Bullards Bar Dam. Extensive HEC-ResSim model development for the Yuba-Feather FIRO Preliminary Viability Assessment (PVA) program has resulted in implementation of three operations alternatives for several event simulations.

Moving forward from the PVA, this document details the simulation plan for the Final Viability Assessment (FVA). Taking into consideration lessons learned from the PVA, the proposed framework will focus on leveraging the PVA and SPK's WCM baseline models, placing greater emphasis on evaluating alternatives with imperfect forecasts, and evaluating spring events.

Implementation overview

An overview of the alternatives being evaluated in the FVA are detailed in Table A-12. An overview of all implementation steps and subsequent actions are detailed in Table A-13.

ID	ORO At-Site	NBB At-Site	System Operation	Description
1	Existing F-CO	Existing F-CO	Existing F-CO	Representation of the current forecast-coordinated operations program. The 120/180 rules will be removed and set for 180k at Yuba City and 180k at Marysville
3	ID3	FIRO Guide Curve	Existing F-CO	Pairing of prescriptive alternatives at dams. The forecast ensemble is processed to a single inflow value (75 percent non-exceedance probability at dam) and is used to determine forecast-based top of conservation or guide curve and/or release magnitude based on pre-defined relationships. Considers forecast duration up to seven days.
4	Hybrid EFO	Hybrid EFO	Existing F-CO	Pairing of iterative alternatives at dams. A potential release from a dam is evaluated considering each forecast ensemble hydrograph. If the tolerable risk of a given outcome, such as exceeding a given reservoir elevation, is exceeded considering the full ensemble, a new release is evaluated. This process is repeated until the tolerable risk is not exceeded. Considers forecast duration up to 15 days.

Table A-12. FVA combined alternatives

Steps	Actions	Responsibility
Metrics	Metrics to be evaluated are detailed in the HEMP	WRE Team
Pre-processing imperfect forecast data	To date, hindcasts are in HEC-DSS, but HDR is working on the configuration into the model	Forecast volumes, NBB GC, downstream constraints: MBK Oroville TOC: HDR EFO release schedules: CW3E
At-site alternative configuration		Oroville: HDR NBB: MBK
Configure FVA HEC-ResSim model		HDR/MBK

Run FVA HEC-ResSim model simulations		HDR/MBK
Post-processing model output		HDR/MBK
Additional sensitivity analysis		HDR

Table A-13. FVA simulation plan implementation summary

Metrics to evaluate FVA alternatives

To compare each alternative, operations with the perfect and imperfect forecasts will be simulated with HEC-ResSim and metrics of success in meeting operation objectives will be computed. These metrics are detailed in the HEMP. The outputs noted in the "FIRO Checklist" document for the July 31 handoff will be developed as well.

Model inputs and set up

Each alternative will be evaluated using a combined HEC-ResSim model. Performance will be evaluated using perfect and imperfect forecasts. A summary of the FVA HEC-ResSim model is detailed in Table A-16.

HEC-ResSim Build	3.5
Starting Point Model Name	YF FIRO BASELINE MODEL.7z
FVA Model Name	YF FIRO FVA MODEL.7z
Workspace Name	YF FIRO FVA MODEL.wksp
Network	Feather-Yuba-Bear_FCO FIRO
Simulations	See Table A-17
Alternatives	See Table A-15 and A-16
Events	See Table A-17

Table A-14. HEC-ResSim details and configuration summary

A new naming convention will be used for the FVA alternatives. Figure A-2 depicts one potential HEC-ResSim simulation alternative naming scheme that would be compatible with all the hydrology, event, and physical configurations. This is meant as a possibility for discussion with the WRE modeling focus team. Alternately, if alternative variants can be assigned clear numeric identifiers that account for structural changes at NBB, the DMP naming scheme can be used.

To describe operations leveraging the hindcast ensembles, each event will be constructed of multiple HEC-ResSim simulations. A simulation will be constructed, executed, and stored for each hindcast issuance date (24-hour interval) using forecast products derived only from current and former forecasts. The simulation on the next date will be initialized with starting storages, releases, and guide curves from the previous simulation. Thus, simulation following a "real time" approach will be used.

"Simulation Set"	Simulation	Lookback	Start	End	Storages, Flows Initialized From	All Forecast Inputs Derived From
1997 x 100%	1996122112Z	18Dec1996 1200	21Dec1996 1200	31Dec1996 1200	Guide Curve	1996122112_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122212Z	19Dec1996 1200	22Dec1996 1200	01Jan1997 1200	1996122112Z	1996122212_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122312Z	20Dec1996 1200	23Dec1996 1200	02Jan1997 1200	1996122212Z	1996122312_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122412Z	21Dec1996 1200	24Dec1996 1200	03Jan1997 1200	1996122312Z	1996122412_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122512Z	22Dec1996 1200	25Dec1996 1200	04Jan1997 1200	1996122412Z	1996122512_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122612Z	23Dec1996 1200	26Dec1996 1200	05Jan1997 1200	1996122512Z	1996122612_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122712Z	24Dec1996 1200	27Dec1996 1200	06Jan1997 1200	1996122612Z	1996122712_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122812Z	25Dec1996 1200	28Dec1996 1200	07Jan1997 1200	1996122712Z	1996122812_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996122912Z	26Dec1996 1200	29Dec1996 1200	08Jan1997 1200	1996122812Z	1996122912_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996123012Z	27Dec1996 1200	30Dec1996 1200	09Jan1997 1200	1996122912Z	1996123012_FeatherYuba_hefs_hourly.csv
1997 x 100%	1996123112Z	28Dec1996 1200	31Dec1996 1200	10Jan1997 1200	1996123012Z	1996123112_FeatherYuba_hefs_hourly.csv

1997 x 100%	1997010112Z	29Dec1996 1200	01Jan1997 1200	11Jan1997 1200	1996123112Z	1997010112_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010212Z	30Dec1996 1200	02Jan1997 1200	12Jan1997 1200	1997010112Z	1997010212_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010312Z	31Dec1996 1200	03Jan1997 1200	13Jan1997 1200	1997010212Z	1997010312_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010412Z	01Jan1997 1200	04Jan1997 1200	14Jan1997 1200	1997010312Z	1997010412_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010512Z	02Jan1997 1200	05Jan1997 1200	15Jan1997 1200	1997010412Z	1997010512_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010612Z	03Jan1997 1200	06Jan1997 1200	16Jan1997 1200	1997010512Z	1997010612_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010712Z	04Jan1997 1200	07Jan1997 1200	17Jan1997 1200	1997010612Z	1997010712_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010812Z	05Jan1997 1200	08Jan1997 1200	18Jan1997 1200	1997010712Z	1997010812_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997010912Z	06Jan1997 1200	09Jan1997 1200	19Jan1997 1200	1997010812Z	1997010912_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997011012Z	07Jan1997 1200	10Jan1997 1200	20Jan1997 1200	1997010912Z	1997011012_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997011112Z	08Jan1997 1200	11Jan1997 1200	21Jan1997 1200	1997011012Z	1997011112_FeatherYuba_hefs_hourly.csv
1997 x 100%	1997011212Z	09Jan1997 1200	12Jan1997 1200	22Jan1997 1200	1997011112Z	1997011212_FeatherYuba_hefs_hourly.csv

Table A-15. Component Simulations of an Example Event Routing with Imperfect Forecasts

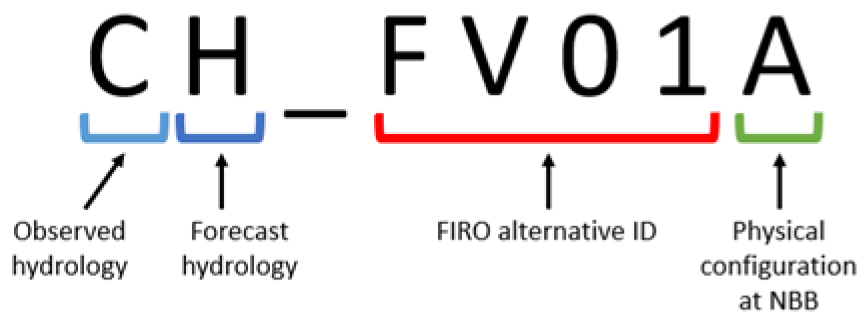


Figure A-2. Example of candidate alternative names for HEC-ResSim modeling

Observed Hydrology and Forecast Hydrology Source IDs:

- C = CVHS
- H = HEFS ensemble or ensemble derivative
- S = HEFS historical simulation
- O = historical observed
- W = WCM hydrology
- SPF1, SPF2, PMF = name of design flood, up to 4 characters (could alternately use "W" as water control manual hydrology ID and the collection member ID 000001 for SPF1, 000002 for SPF2, as in DMP)

Physical Configuration IDs:

- A = ARC spillway included
- E = without ARC spillway (existing physical condition)

Alternative Name	Observed Hydrology	Forecast Hydrology	Applicable Patterns	Applicable Scale Groups (%)
CC_FV**A	CVHS	CVHS	1956, 1965, 1986, 1997, 2006, 2017	10 - 340
SS_FV**A	HEFS2023 historical simulation	HEFS2023 historical simulation	1986, 1995, 1997	1986: 100 - 150 1997: 84 - 130 1995: 50 - 190
SH_FV**A	HEFS2023 historical simulation	HEFS2023 forecasts	1986, 1995, 1997	1986: 100 - 150 1997: 84 - 130 1995: 50 - 190

CH_FV**A	CVHS	HEFS2023	1986, 1997, 2006, 2017	100

Table A-16. FVA HEC-ResSim alternatives names for consolidated model (winter simulations)

Alternative Name	Oroville Operations Set	NBB Operations Set	ARC Spillway	System Operations	Alternative ID	Network
**_FV01A	F-CO EventOp	ID1A	With ARC Spillway	Explicit system Balance	ID1	Feather-Yuba-Bear_FCO FIRO
**_FV03A	ID3	ID3A	With ARC Spillway	Explicit system Balance	ID3	Feather-Yuba-Bear_FCO FIRO
**_FV04A	ID4	ID4A	With ARC Spillway	Explicit system Balance	ID4	Feather-Yuba-Bear_FCO FIRO
**_FV01E	F-CO_FIRO Baseline	F-CO_FIRO Baseline	Without ARC Spillway	Explicit system Balance	ID1	Feather-Yuba-Bear_FCO FIRO
**_FV03E	ID3	ID3E	Without ARC Spillway	Explicit system Balance	ID3	Feather-Yuba-Bear_FCO FIRO
**_FV04E	ID4	ID4E	Without ARC Spillway	Explicit system Balance	ID4	Feather-Yuba-Bear_FCO FIRO

Table A-17. FVA HEC-ResSim alternative names for representing with and without ARC Spillway configurations

Simulation Name	Dataset	Event	Scale Groups (%)	Hydrology Time Zone	Simulation Lookback	Simulation Start	Simulation End	FVA Requirement	WCM Requirement
1956	CVHS	1956	10 – 340	Pacific	12 Dec 1955, 0000	15 Dec 1955, 0000	12 Jan 1956, 0000		X

1965	CVHS	1965	10 – 340	Pacific	12 Dec 1964, 0000	15 Dec 1964, 0000	12 Jan 1965, 0000		X
1986	CVHS	1986	10 – 340	Pacific	04 Feb 1986, 0000	06 Feb 1986, 0000	14 Mar 1986, 0400		X
1997	CVHS	1997	10 – 340	Pacific	18 Dec 1996, 0000	21 Dec 1996, 0000	12 Jan 1997, 0000		X
2006	CVHS	2006	100	Pacific	06 Dec 2005, 0400	08 Dec 2005, 0400	09 Jan 2006, 0000		
2017	CVHS	2017	100	Pacific	28 Jan 2017, 0400	01 Feb 2017, 0400	27 Feb 2017, 0400		
1986	GEFSv12 hindcasts	1986	100, 102, 104, 106, 108, 110, 120, 130, 140, 150	UTC	04 Feb 1986, 1200	06 Feb 1986, 1200	14 Mar 1986, 1200	X	X
1997	GEFSv12 hindcasts	1997	84, 86, 88, 90, 92, 94, 96, 98, 100, 110, 120, 130	UTC	18 Dec 1996, 1200	21 Dec 1996, 1200	12 Jan 1997, 1200	X	X
Mar1995_1	GEFSv12 hindcasts	Mar1995	50, 70, 90, 100, 110, 130	UTC	27 Feb 1995, 1200	1 Mar 1995, 1200	16 Mar 1995, 1200	X	
Mar1995_2	GEFSv12 hindcasts	Mar1995′	50, 70, 90, 100, 110, 130	UTC	29 Mar 1995, 1200	31 Mar 1995, 1200	14 Apr 1995, 1200	X	
Mar1995_3	GEFSv12 hindcasts	Mar1995″	50, 70, 90, 100, 110, 130	UTC	12 Apr 1995, 1200	14 Apr 1995, 1200	29 Apr 1995, 1200	X	
May1995_1	GEFSv12 hindcasts	May1995	70, 90, 100, 110, 130, 150, 170, 190	UTC	23 Apr 1995, 1200	25 Apr 1995, 1200	7 May 1995, 1200	X	
May1995_2	GEFSv12 hindcasts	May1995′	70, 90, 100, 110, 130, 150, 170, 190	UTC	13 May 1995, 1200	15 May 1995, 1200	27 May 1995, 1200	X	

Table A-18. FVA HEC-ResSim simulations

File Description	Notes
Dummy flows (timeseries of 0 values as placeholders for modeling with different hydrologic boundary conditions)	This will facilitate switching between CVHS and CNRFC model boundary conditions within the same model network
CVHS Scaled Simulated Historical Hydrology (2023 update for Yuba-Feather CVHS hydrology)	Use latest CVHS hydrology (extended during update to annual flow-frequency curves)
CNRFC HEFS 2023 scaled historical simulation hydrology	Use latest scaled simulations; derivatives of the ensemble forecasts for each forecast date and scaled event feed into the following rows
Oroville Inputs derived from perfect forecast – 1-day average inflows – 3-day average inflows – 5-day average inflows – 7-day average inflows – FIRO top of conservation timeseries – ORO EFO release schedule	All information needed to specify Oroville at-site FIRO releases based on ID3 or ID4 for both hydrology datasets
Oroville Inputs derived from imperfect forecasts – 1-day average inflows (75 % NEP) – 3-day average inflows (75 % NEP) – 5-day average inflows (75 % NEP) – 7-day average inflows (75 % NEP) – FIRO top of conservation timeseries – ORO EFO release schedule	All information needed to specify Oroville at-site FIRO releases based on ID3 or ID4 that are derived from the ensemble hindcasts
NBB Inputs derived from perfect forecast – FIRO guide curve storage timeseries – NBB EFO release schedule	All information needed to specify NBB at-site FIRO releases based on ID3 or ID4 for both hydrology datasets
NBB Inputs derived from imperfect forecasts – FIRO guide curve storage timeseries – NBB EFO release schedule	All information needed to specify NBB at-site FIRO releases based on ID3 or ID4 that are derived from the ensemble hindcasts
Flow Space Inputs derived from perfect forecast – difference between Feather at Yuba City constraint and forecast local flows – difference between Feather below Yuba constraint and forecast local flows – difference between Feather below Bear constraint and forecast local flows – difference between Yuba at Marysville constraint and forecast local flows	This may not be necessary, but would be a useful check in developing the technique for defining downstream control flow rules as a function of an external variable

Flow Space Inputs derived from imperfect forecast – difference between Feather at Yuba City constraint and forecast local flows – difference between Feather below Yuba constraint and forecast local flows – difference between Feather below Bear constraint and forecast local flows – difference between Yuba at Marysville constraint and forecast local flows	Stores the "external variable" timeseries that incorporates a representation of the imperfect forecast and the maximum objective flow at the control location to integrate imperfect flow forecast info into the release decisions made for downstream constraints.
WCM Events / PMF (if we determine these events need to be added)	Flow boundary conditions for events like the SPF and PMF

Table A-19. Types of FVA Hydrologic Inputs

File Name from FVA	FIRO Alternatives	Patterns
DUMMY_flows.dss	All	All (unscaled)
NBB_FIRO_space_bounds.dss	ID3E, ID3A, ID4E, ID4A	All (does not change between magnitudes)
CVHS_NBB_perfect_forecast_guide_curve.dss	ID3E, ID3A	1956, 1965, 1986, 1997, 2006, 2017
HEFS_NBB_perfect_forecast_guide_curve_UTC.dss	ID3E, ID3A	1986, 1997, Mar1995, May1995
HEFS_NBB_75NEP_forecast_guide_curve_UTC.dss	ID3E, ID3A	1986, 1997, Mar1995, May1995
CVHS_FeatherYubaBear_scaled_records_DSS7.dss	All	1956, 1965, 1986, 1997

Table A-20. Inventory of FVA Hydrologic Inputs Stored in HEC-ResSim "shared" Directory

File Name from FVA	FIRO Alternatives	Patterns
DUMMY_flows.dss	All	All (unscaled)
NBB_FIRO_space_bounds.dss	ID3E, ID3A, ID4E, ID4A	All (does not change between magnitudes)
CVHS_NBB_perfect_forecast_guide_curve.dss	ID3E, ID3A	1956, 1965, 1986, 1997, 2006, 2017

HEFS_NBB_perfect_forecast_guide_curve_UTC.dss	ID3E, ID3A	1986, 1997, Mar1995, May1995
HEFS_NBB_75NEP_forecast_guide_curve_UTC.dss	ID3E, ID3A	1986, 1997, Mar1995, May1995
CVHS_FeatherYubaBear_scaled_records_DSS7.dss	All	1956, 1965, 1986, 1997

Table A-21. Staring conditions

A.4 Scalings of the 1986 Flood Event

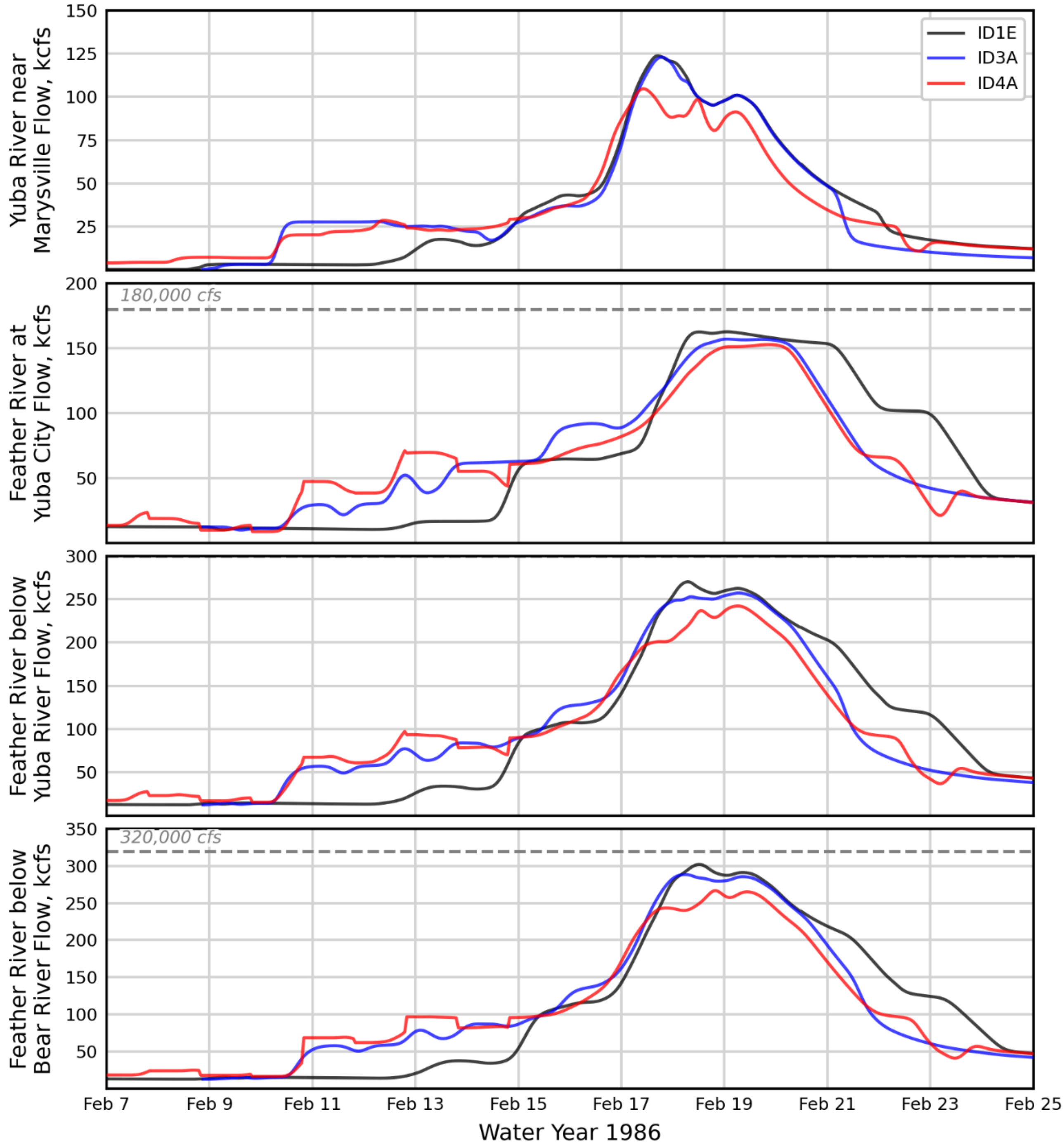
The scaling process is described in Section 5.3.2.2

Scale factors applied: 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 130, 140

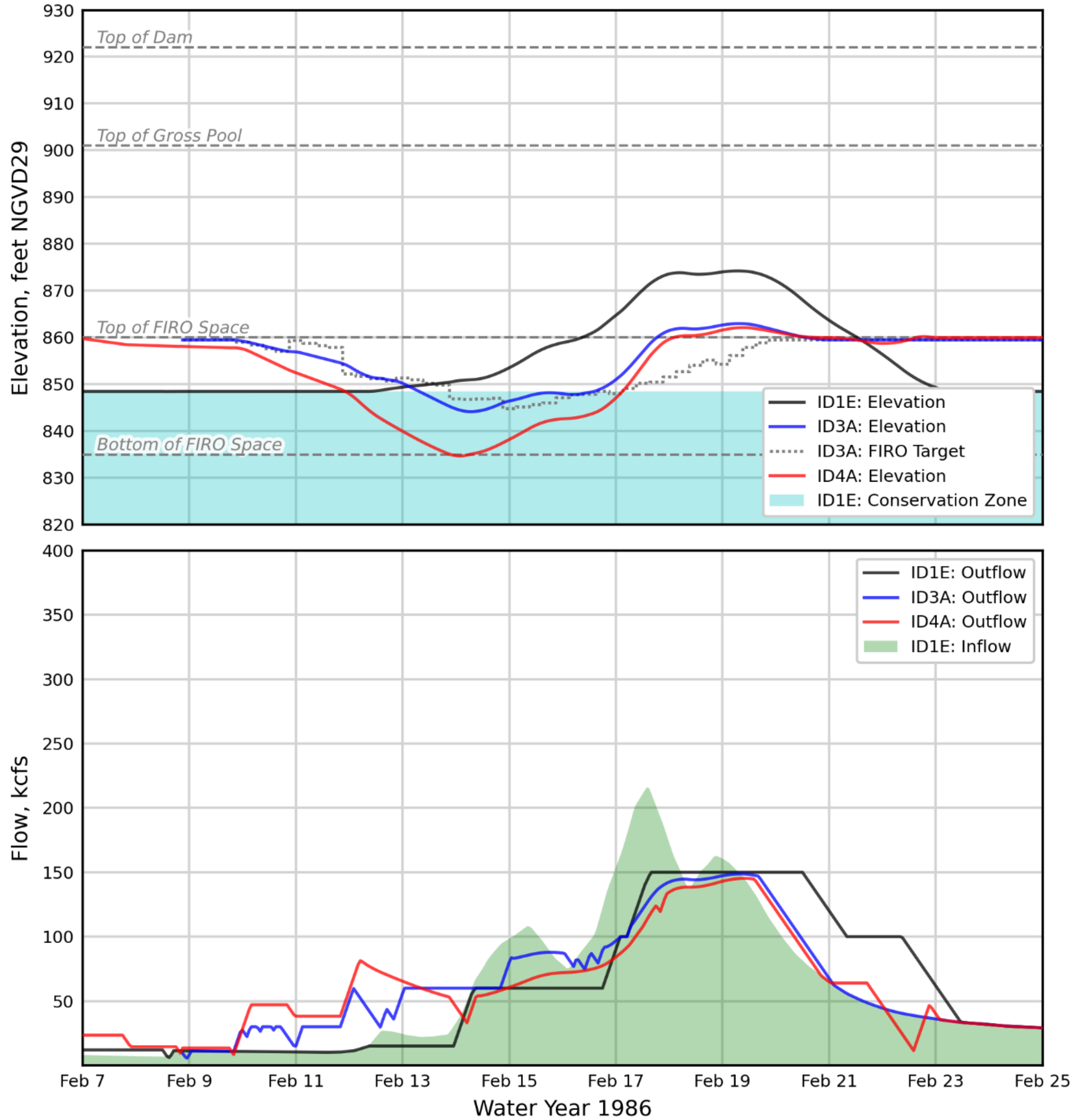
Each scaling contains:
- A plot of the four downstream control points
- A plot of Oroville elevation (top) and releases (bottom) with inflow as the background

Alternatives (ID1E, ID3A, and ID4A) are described in Section 3.

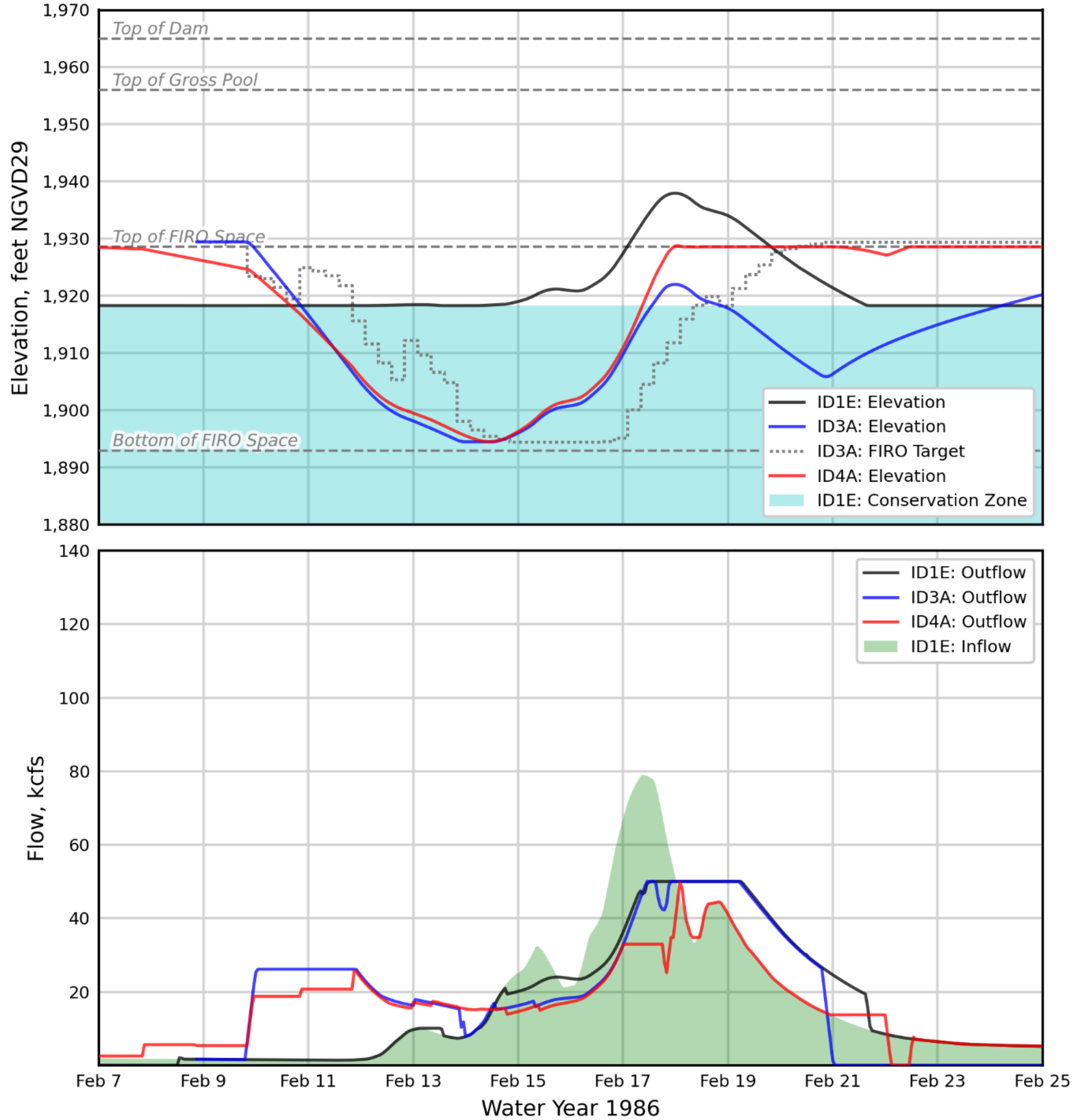
1986 100% Scaling
Downstream Control Points



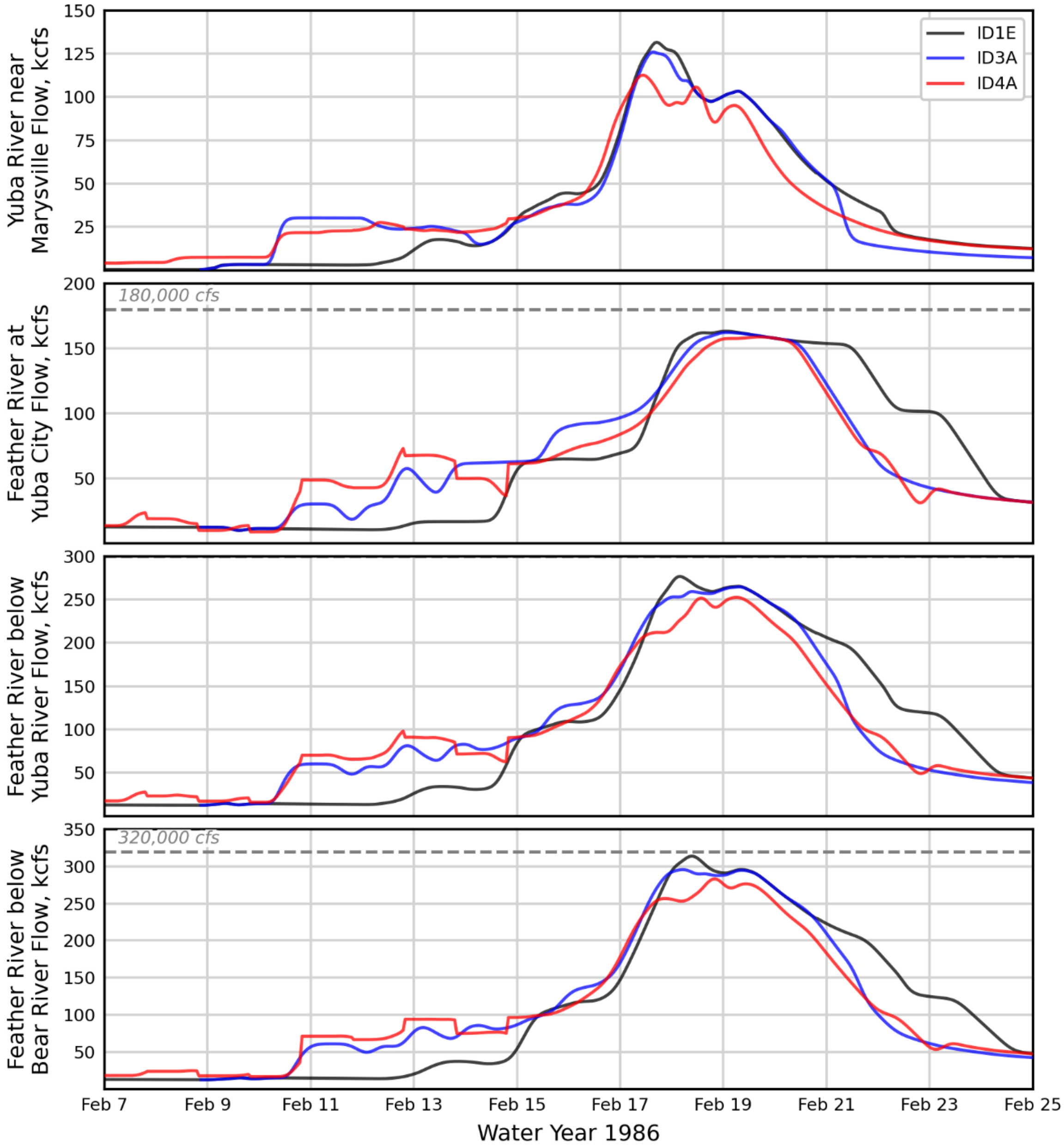
1986 100% Scaling Oroville



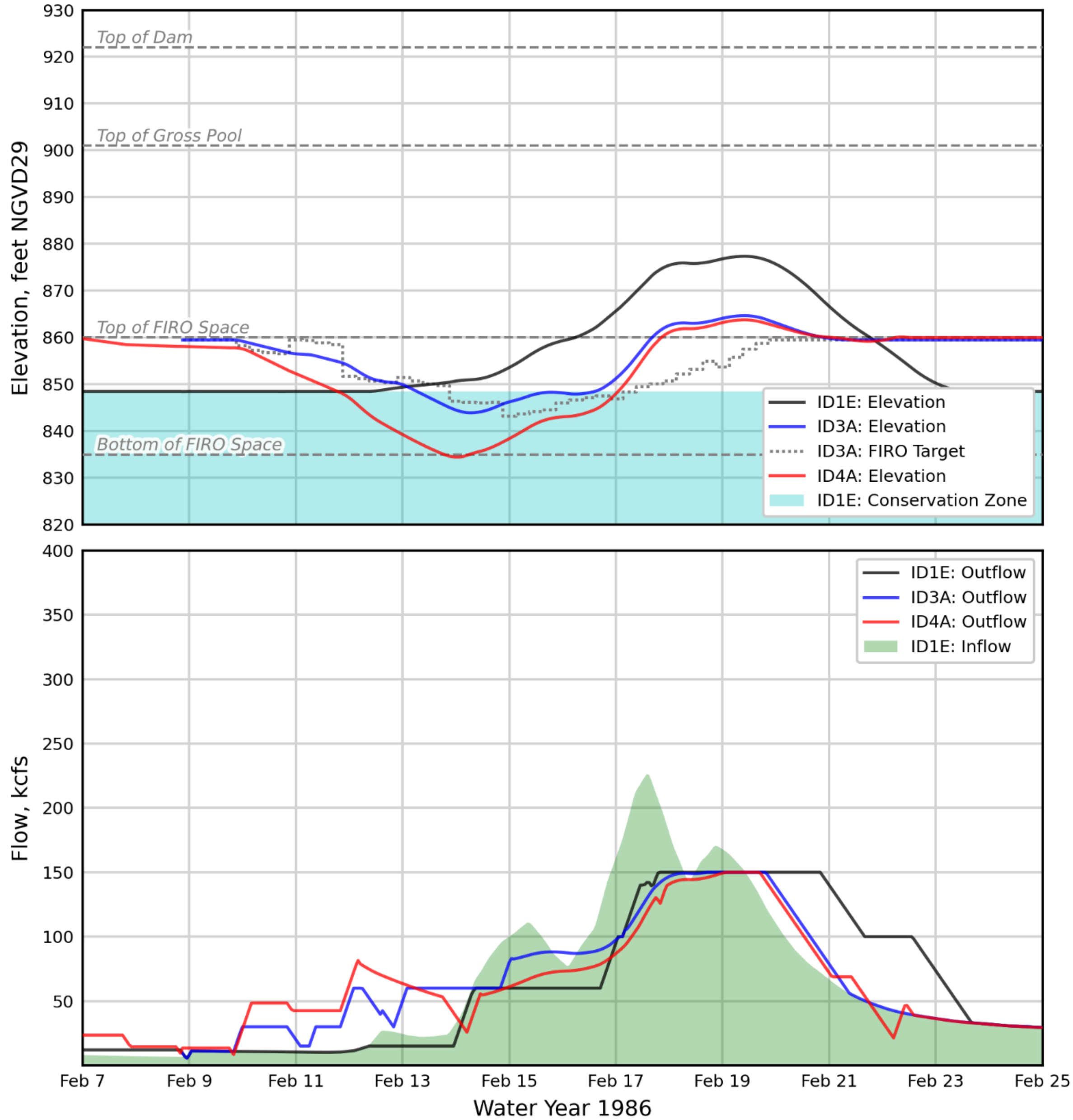
1986 100% Scaling
New Bullards Bar



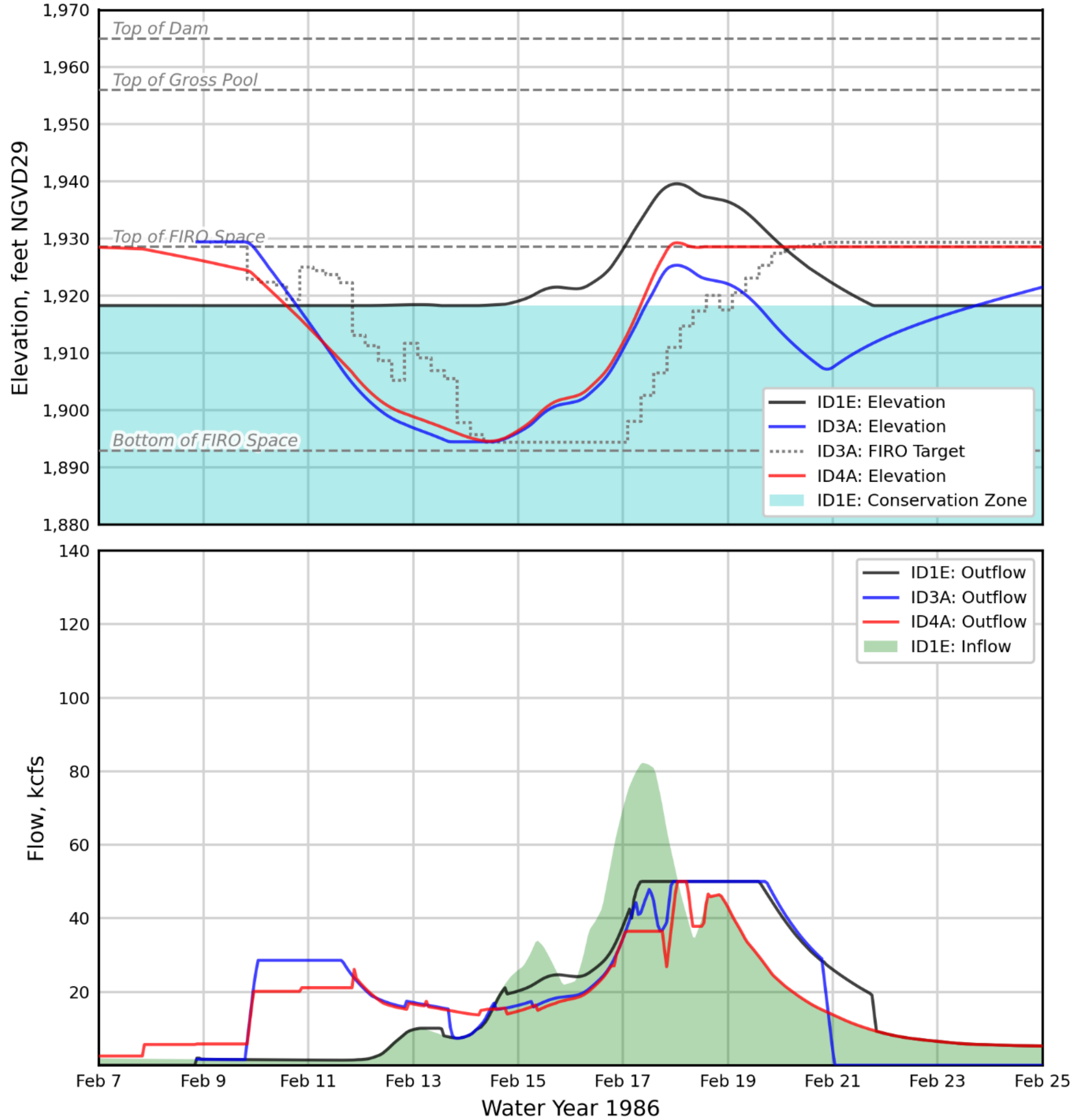
1986 102% Scaling
Downstream Control Points



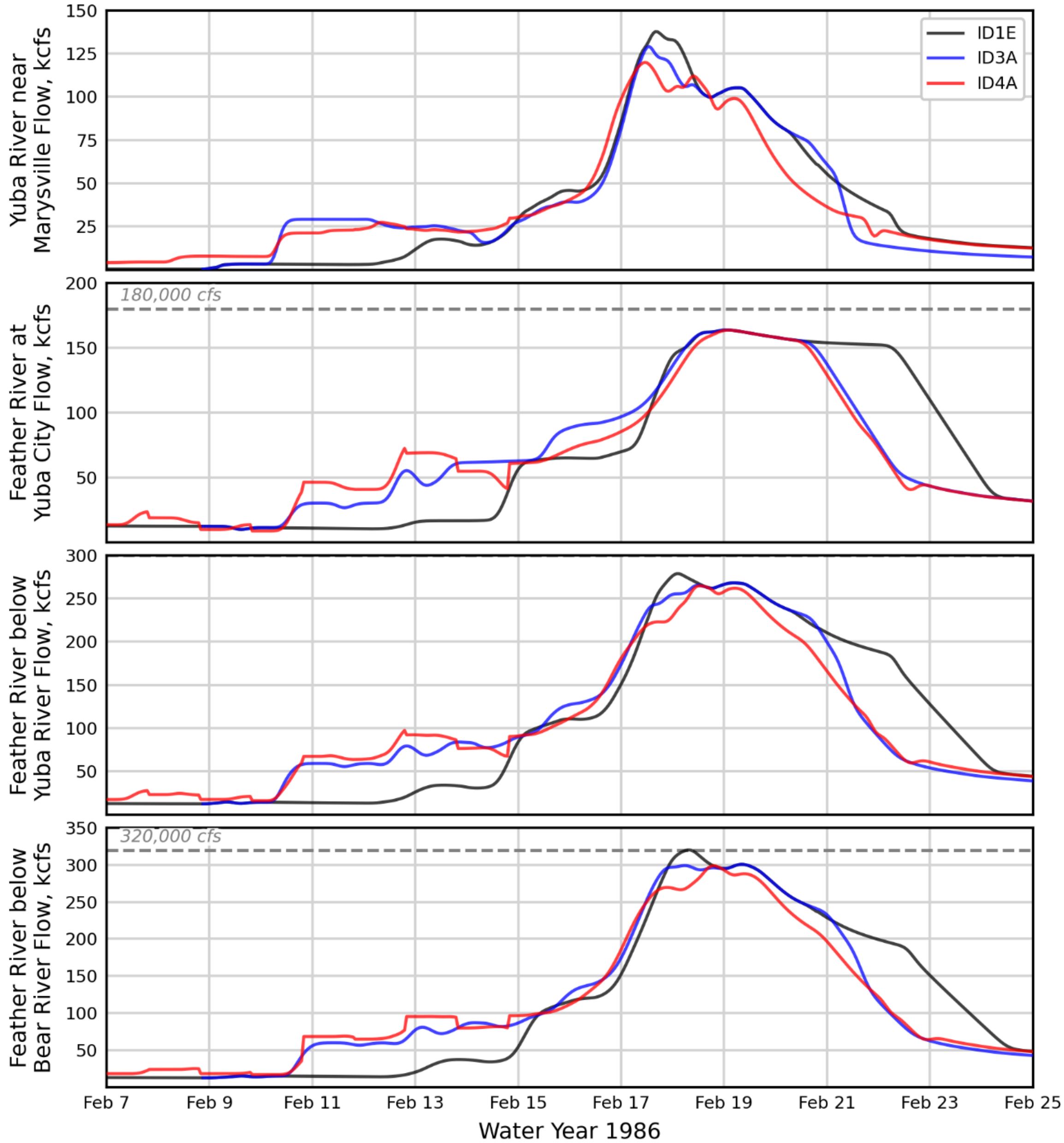
1986 102% Scaling
Oroville



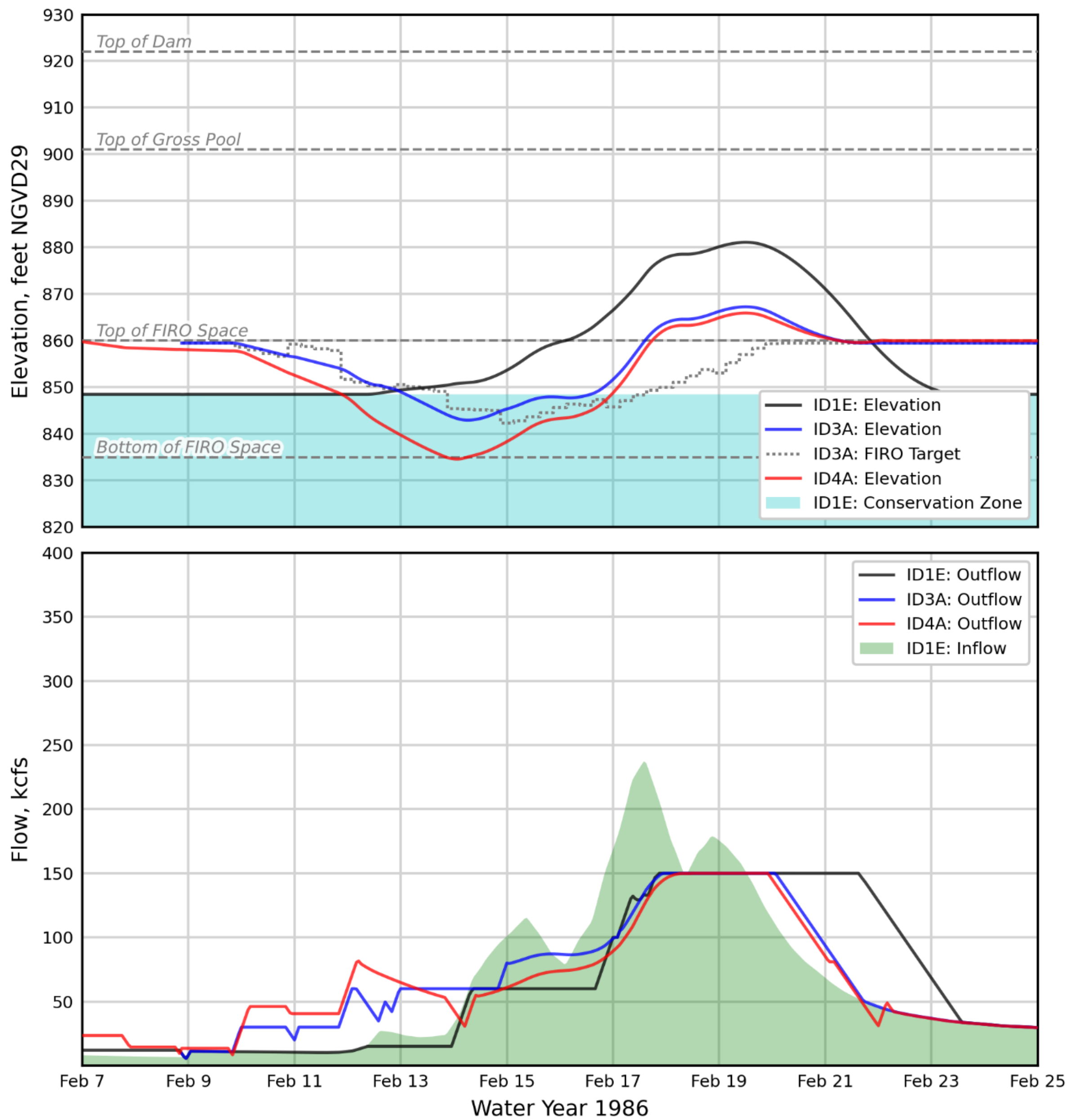
1986 102% Scaling
New Bullards Bar



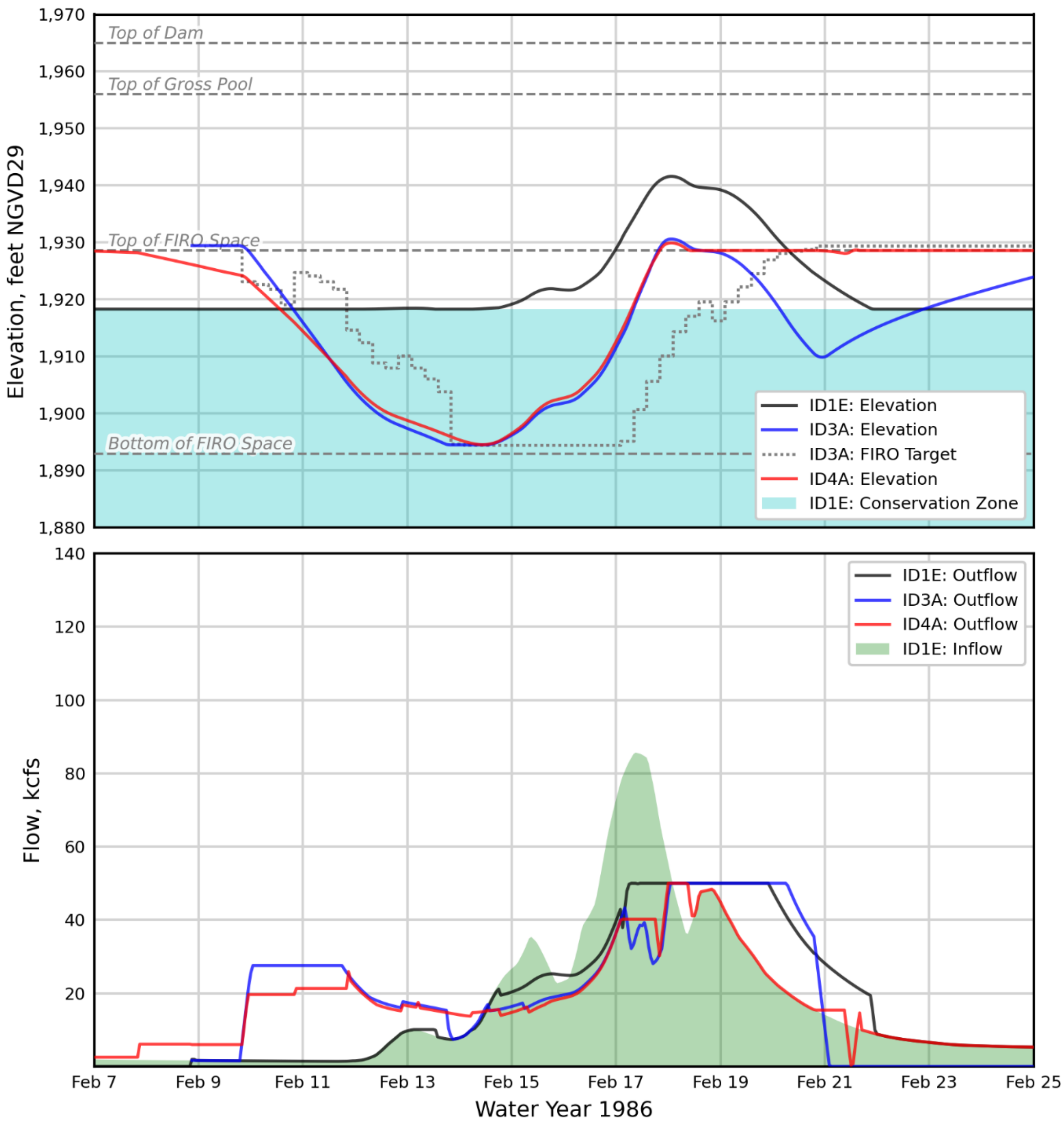
1986 104% Scaling
Downstream Control Points



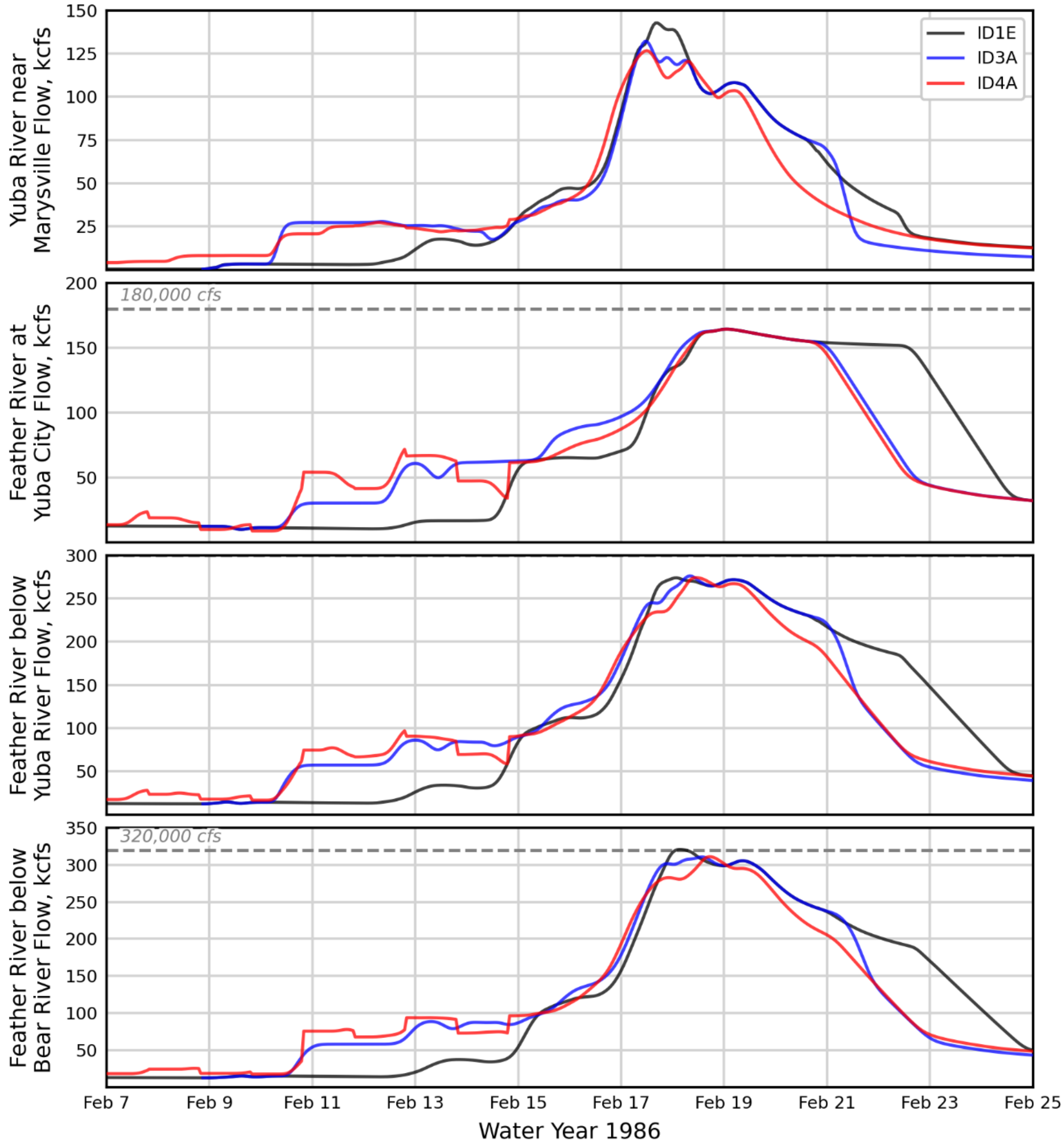
1986 104% Scaling
Oroville



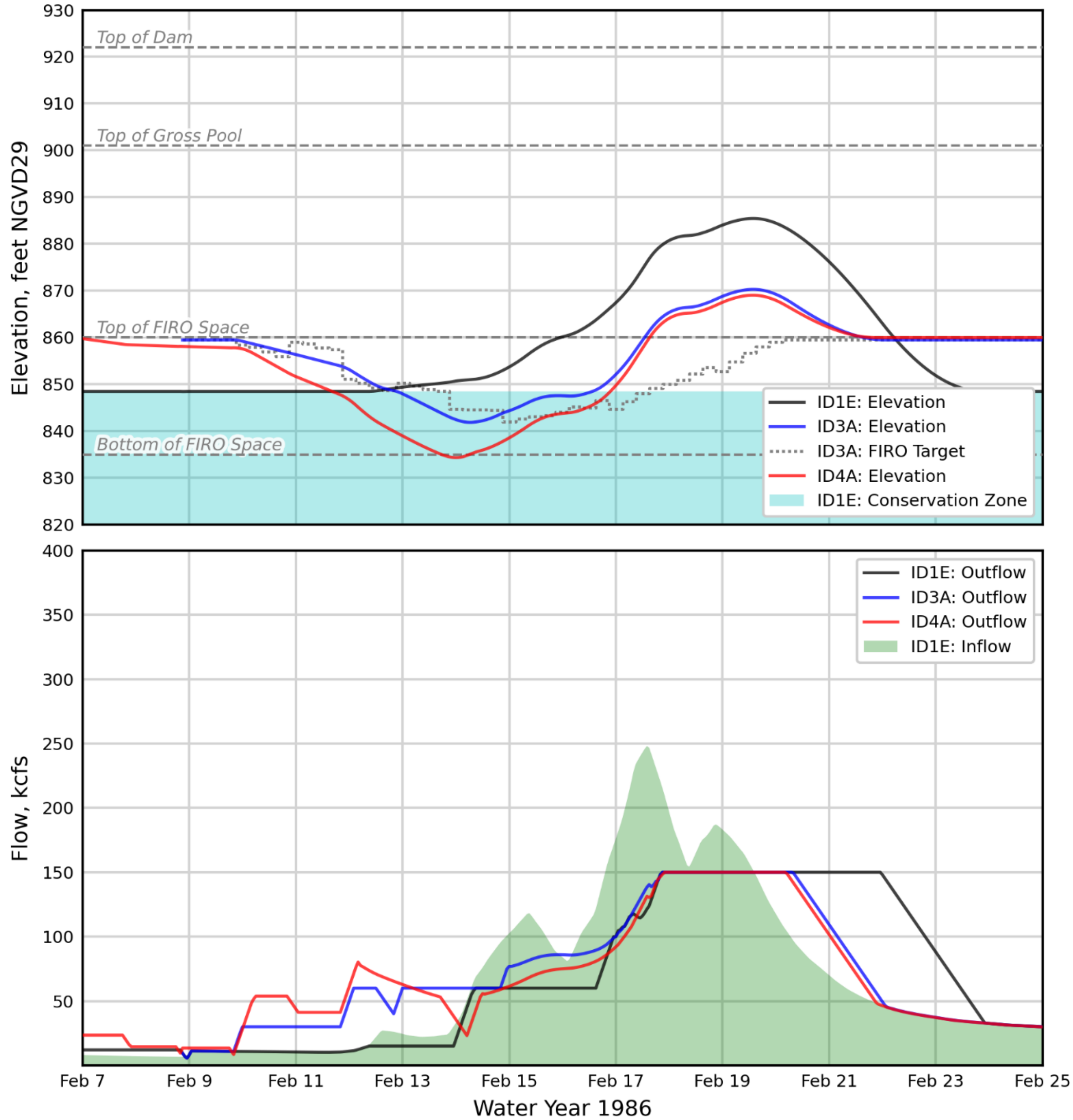
1986 104% Scaling
New Bullards Bar



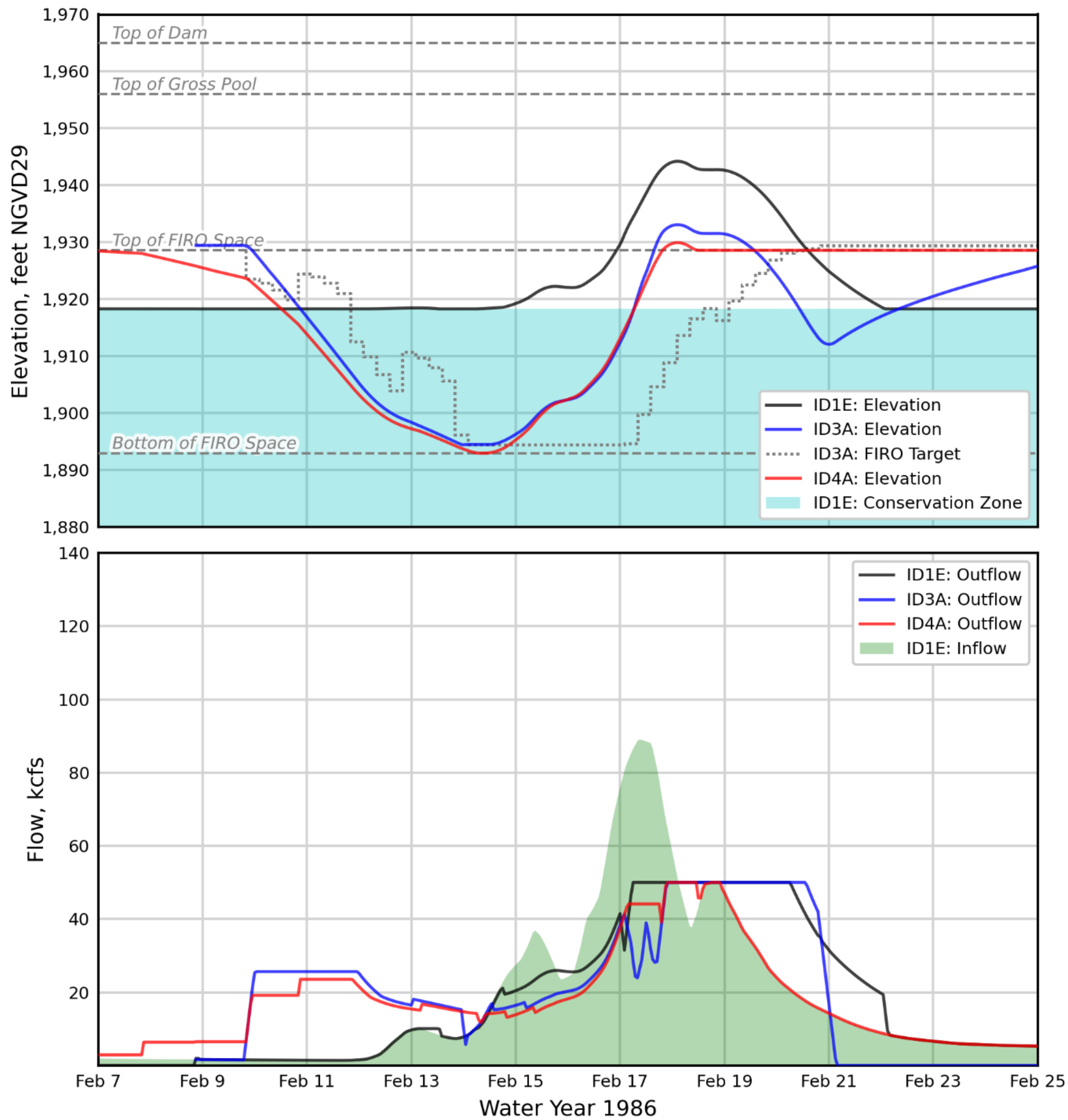
1986 106% Scaling
Downstream Control Points



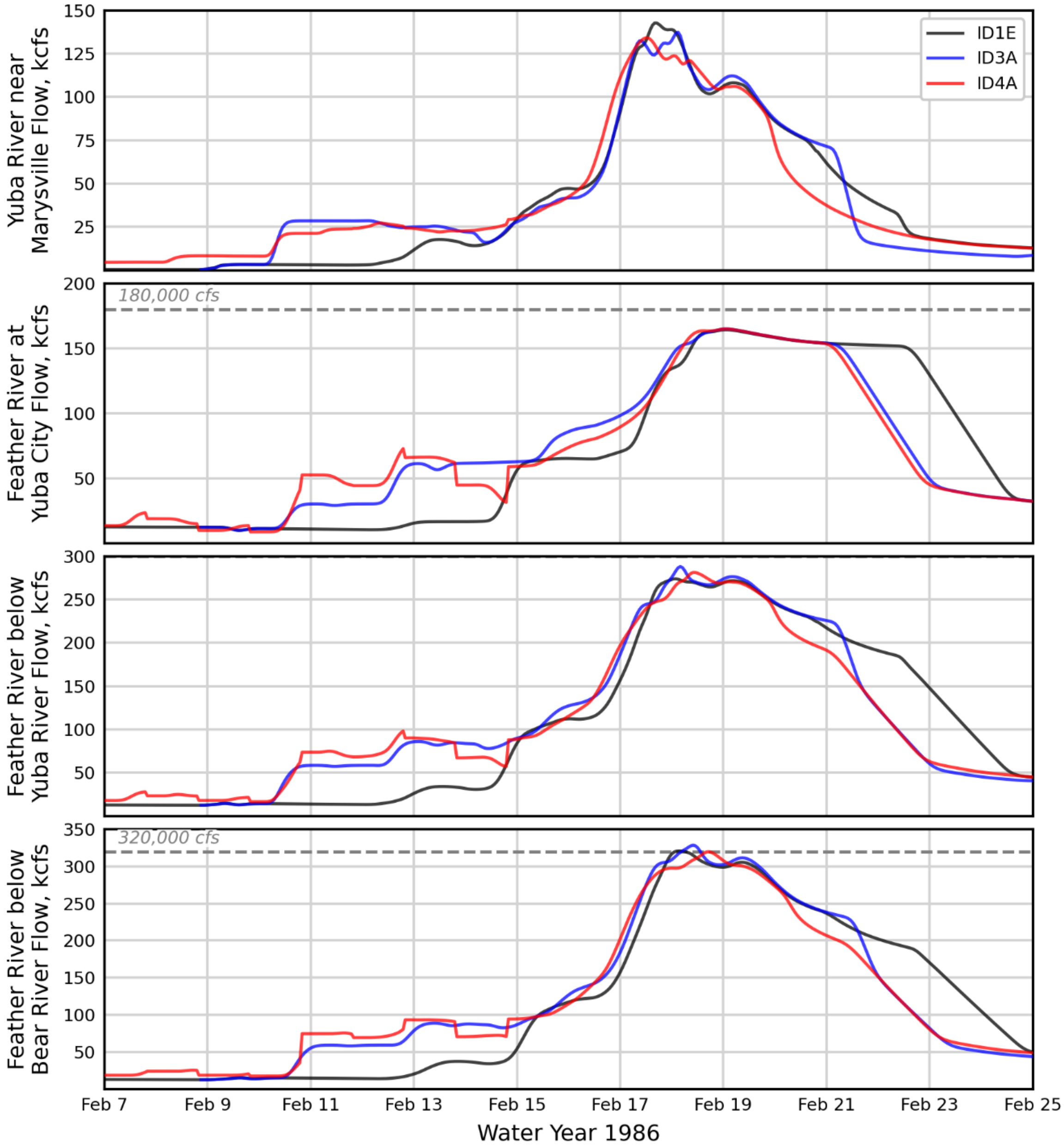
1986 106% Scaling Oroville



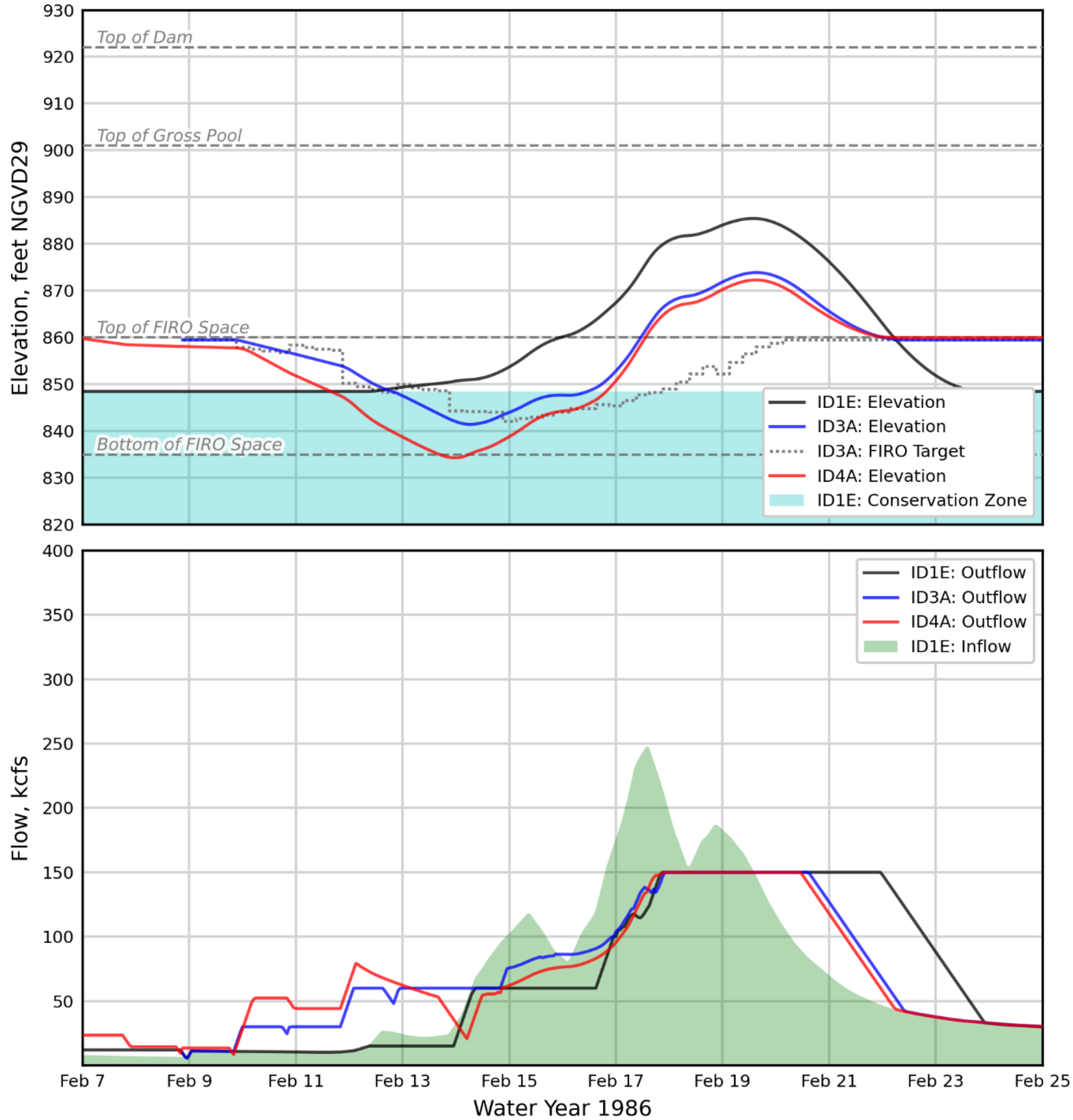
1986 106% Scaling
New Bullards Bar



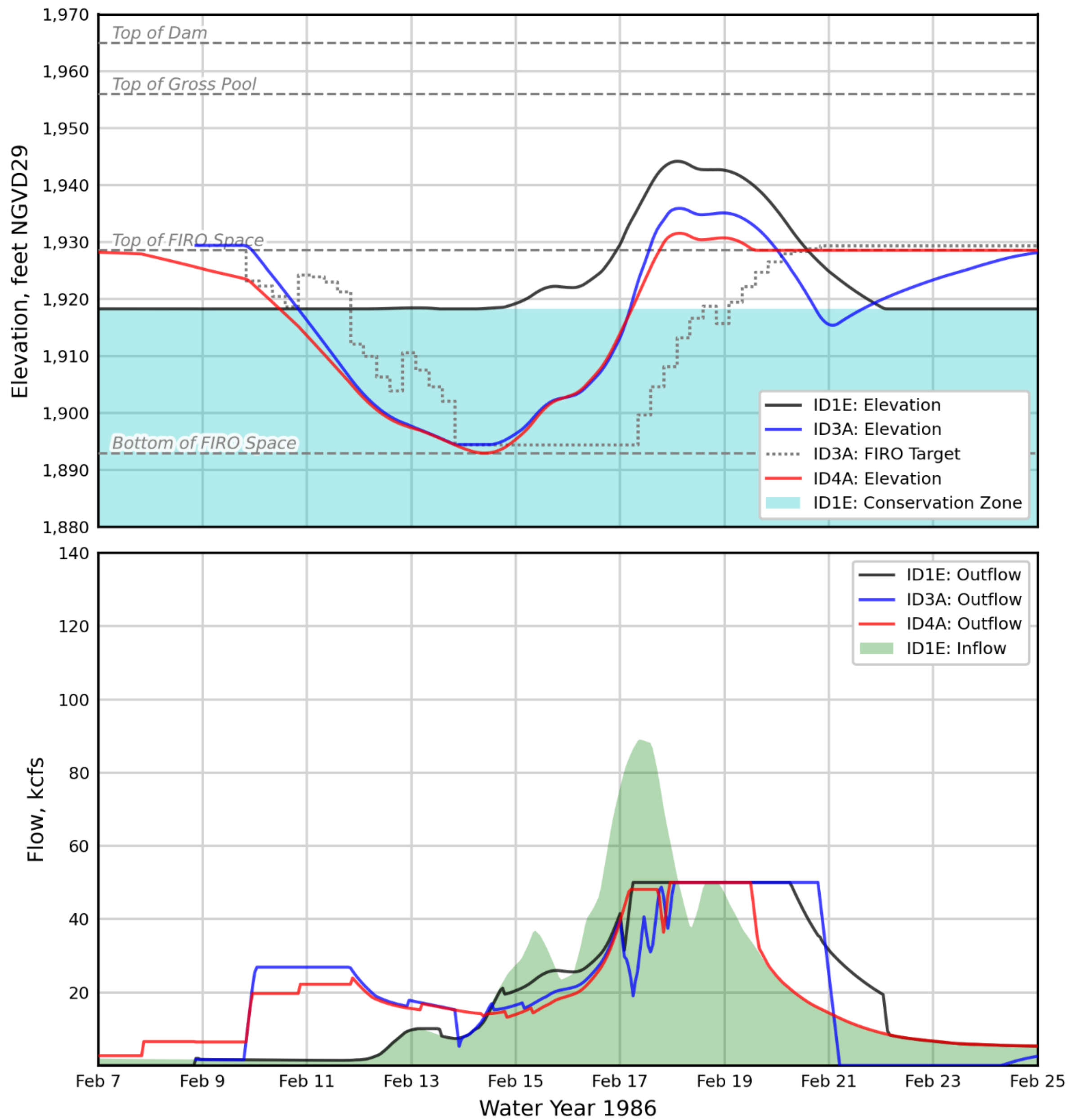
1986 108% Scaling
Downstream Control Points



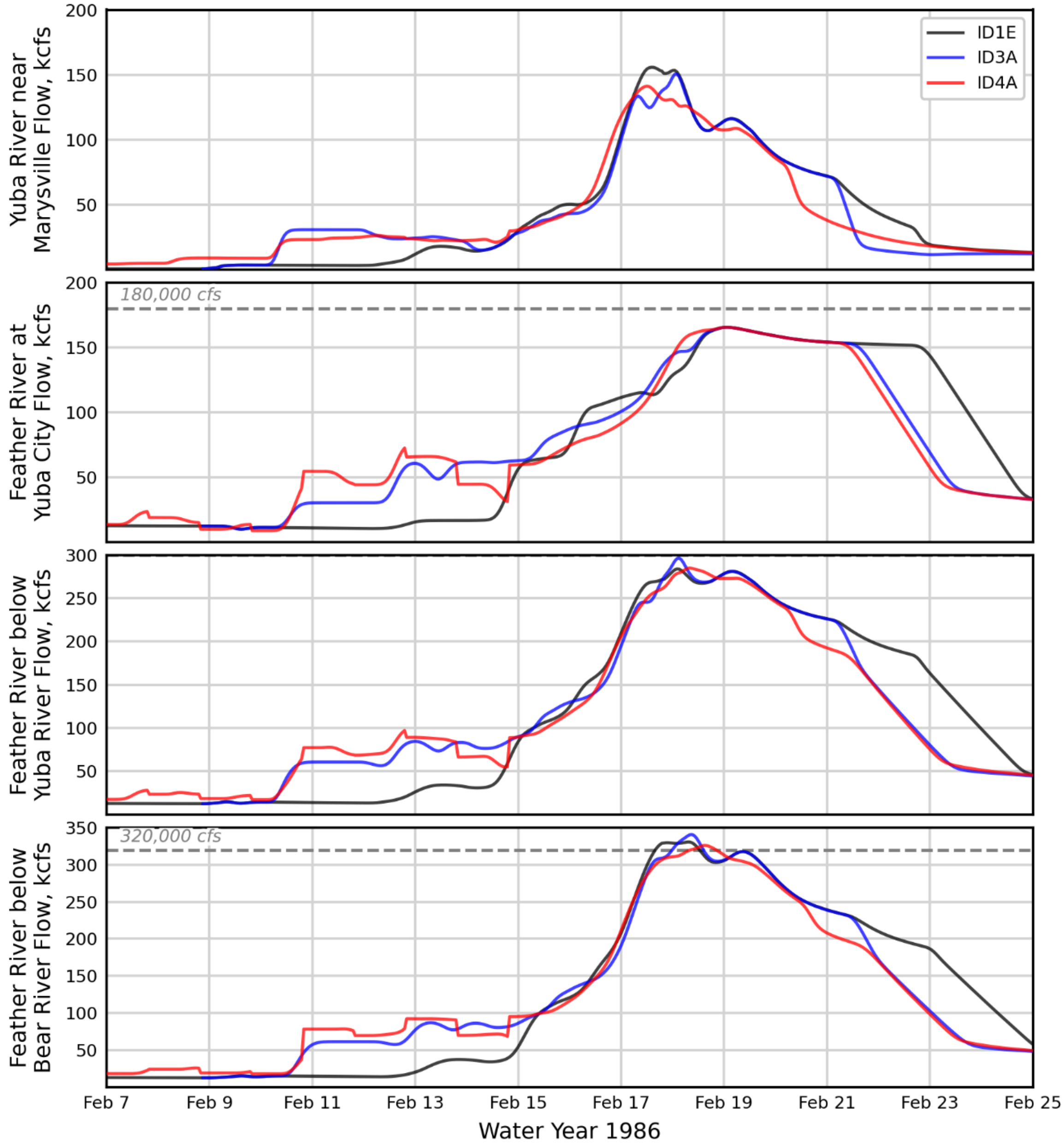
1986 108% Scaling
Oroville



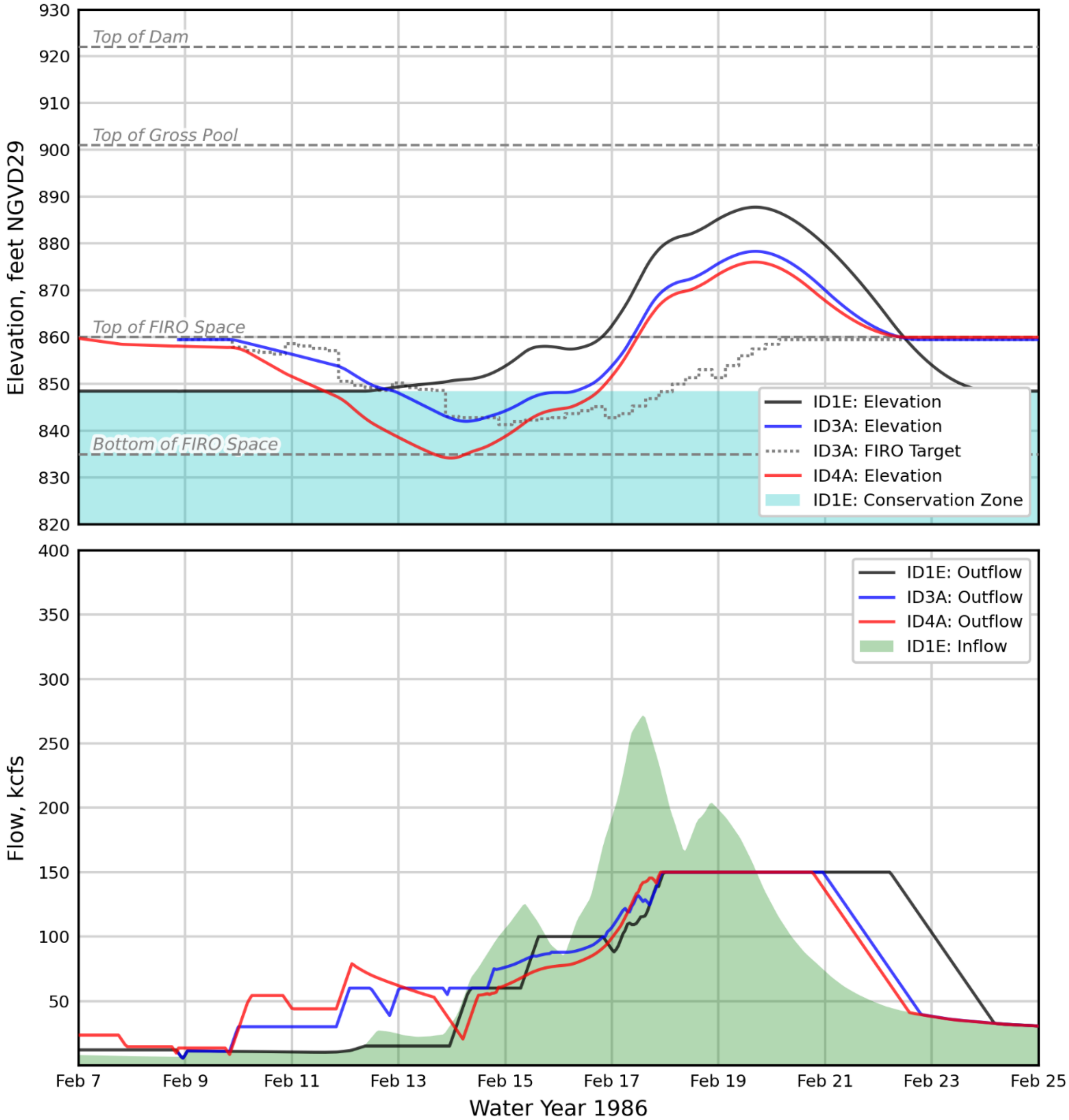
1986 108% Scaling
New Bullards Bar


Top of Dam
Top of Gross Pool
Top of FIRO Space
Bottom of FIRO Space
Elevation, feet NGVD29
1,970
1,960
1,950
1,940
1,930
1,920
1,910
1,900
1,890
1,880
ID1E: Elevation
ID3A: Elevation
ID3A: FIRO Target
ID4A: Elevation
ID1E: Conservation Zone
Flow, kcfs
140
120
100
80
60
40
20
ID1E: Outflow
ID3A: Outflow
ID4A: Outflow
ID1E: Inflow
Feb 7
Feb 9
Feb 11
Feb 13
Feb 15
Feb 17
Feb 19
Feb 21
Feb 23
Feb 25
Water Year 1986

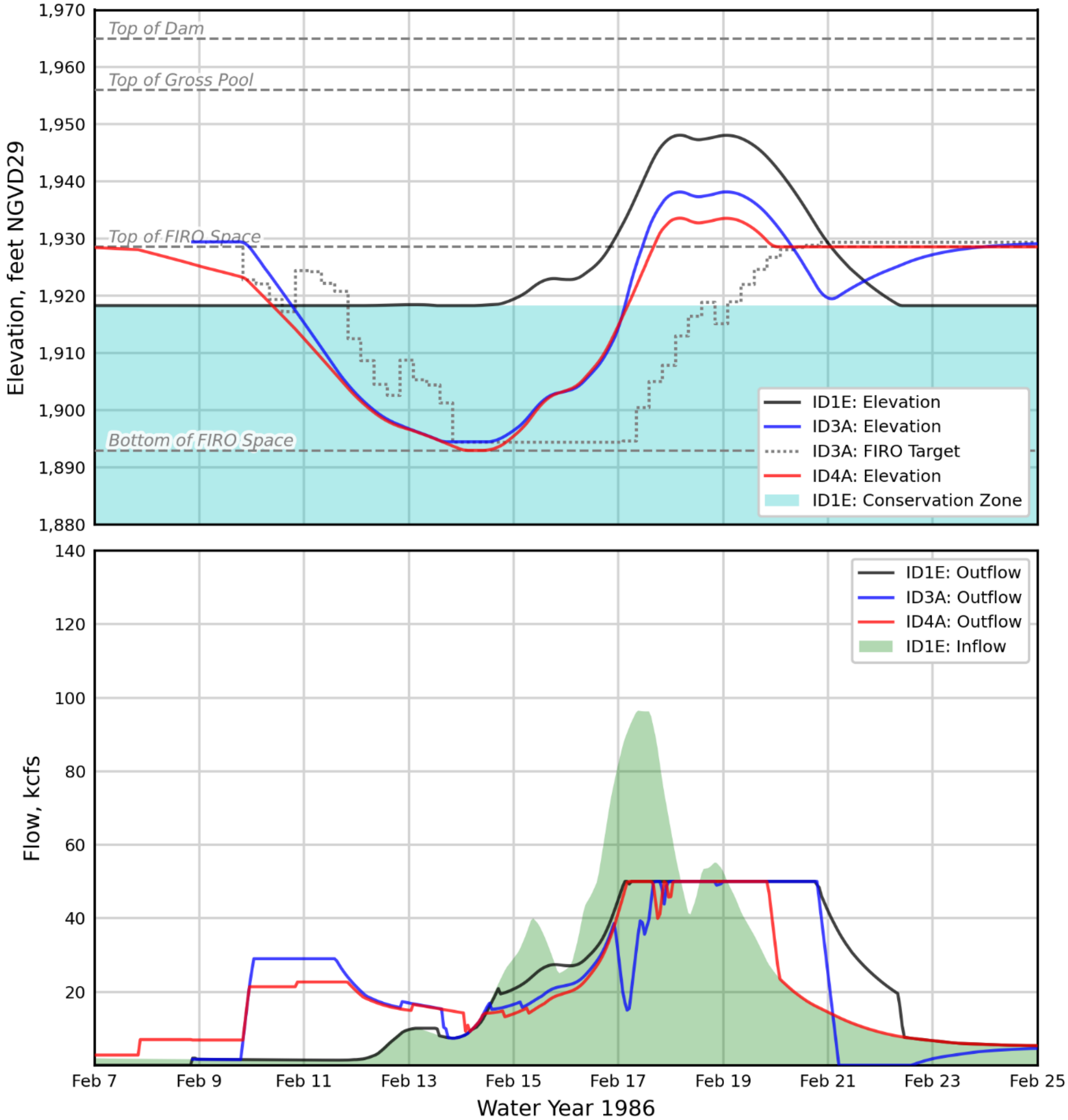
1986 110% Scaling
Downstream Control Points



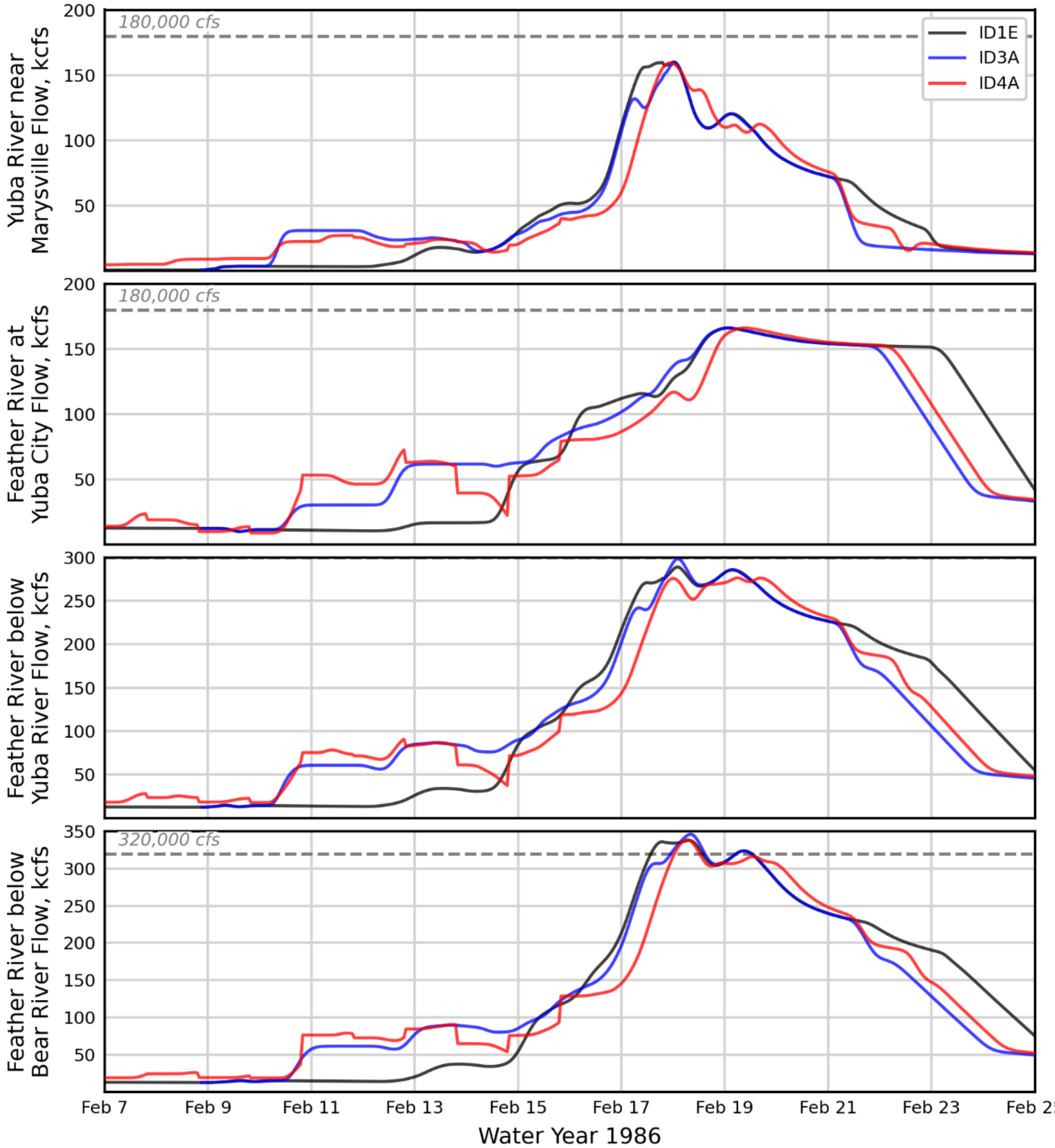
1986 110% Scaling
Oroville



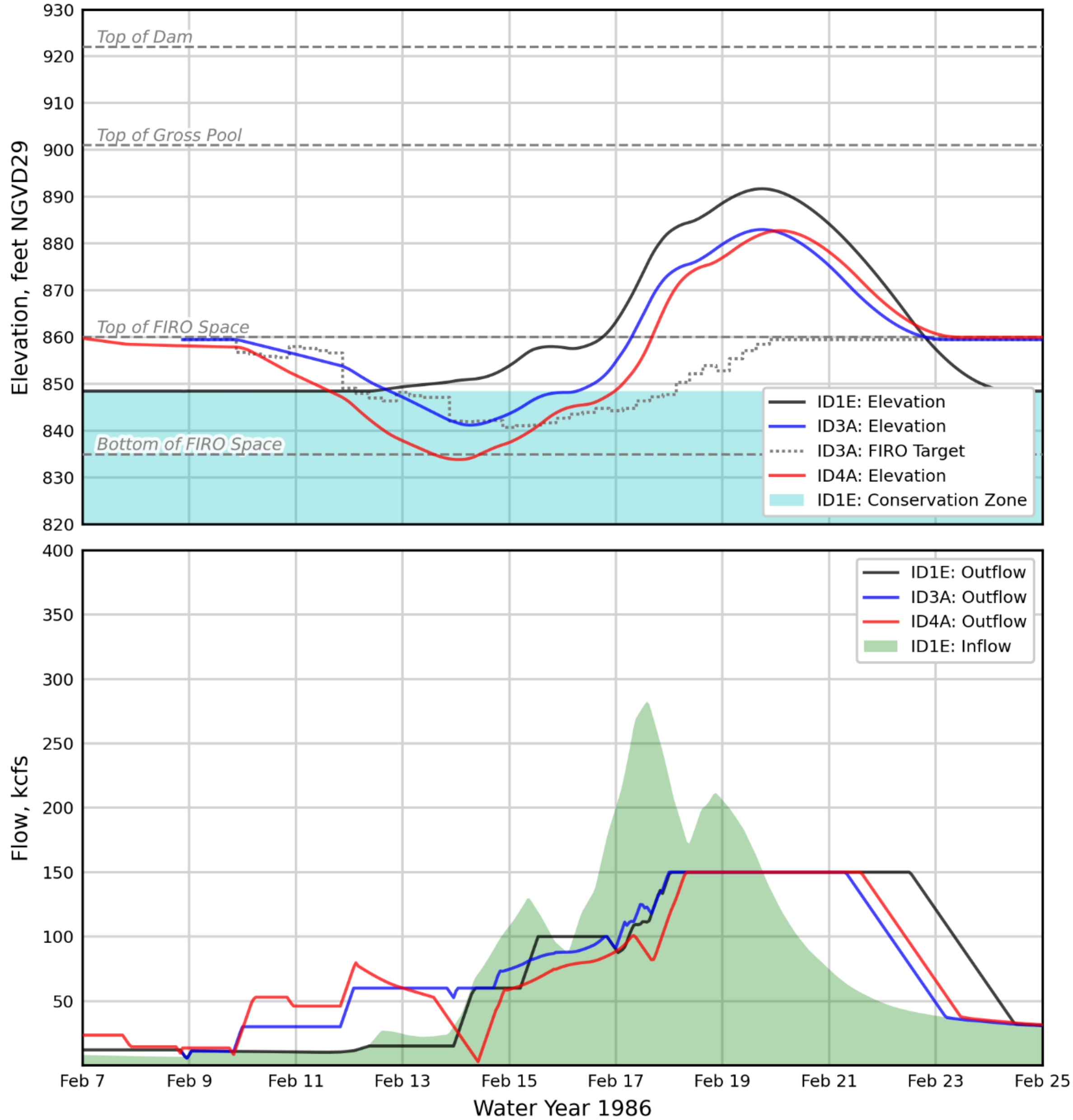
1986 110% Scaling
New Bullards Bar



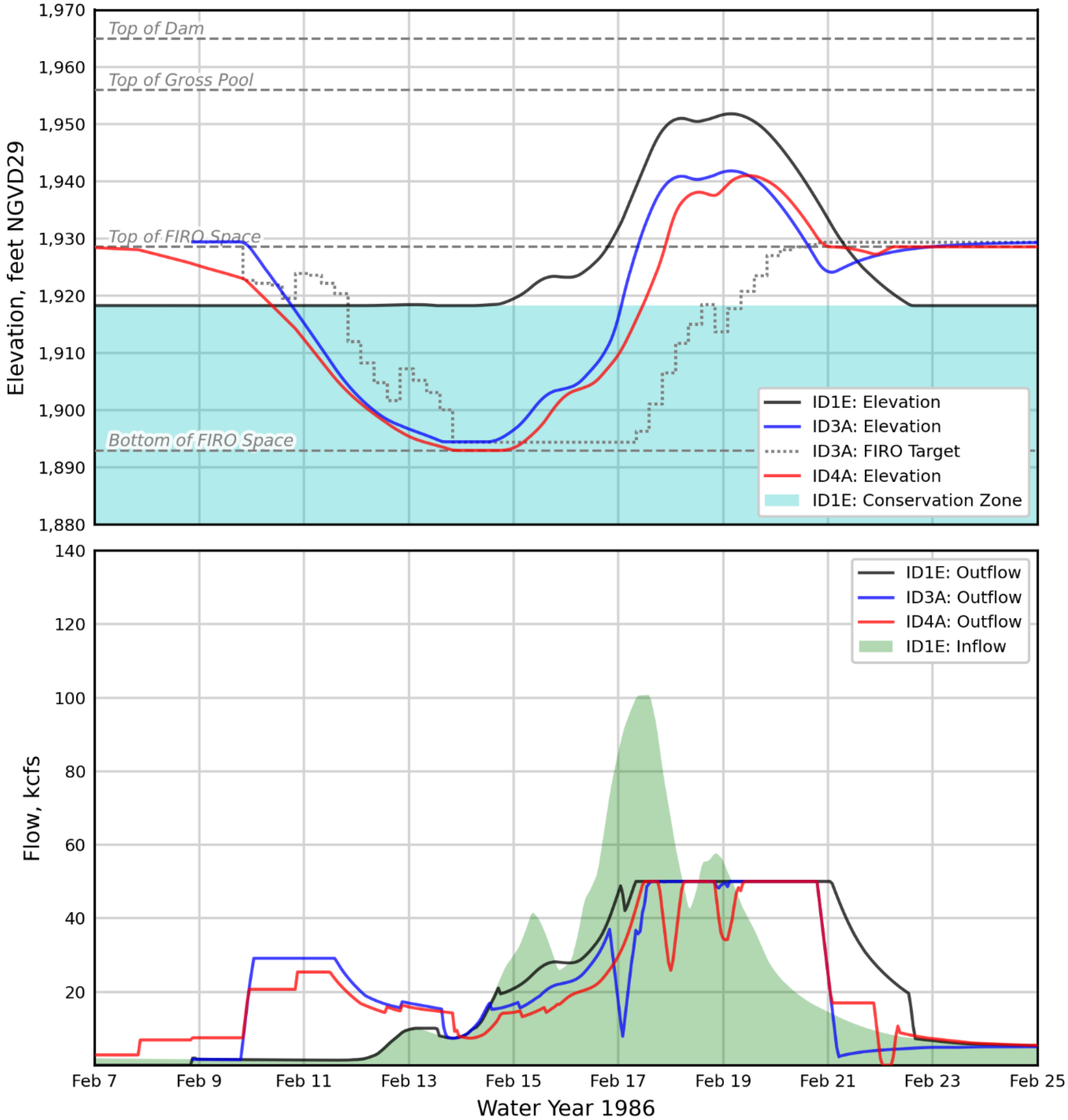
1986 112% Scaling
Downstream Control Points



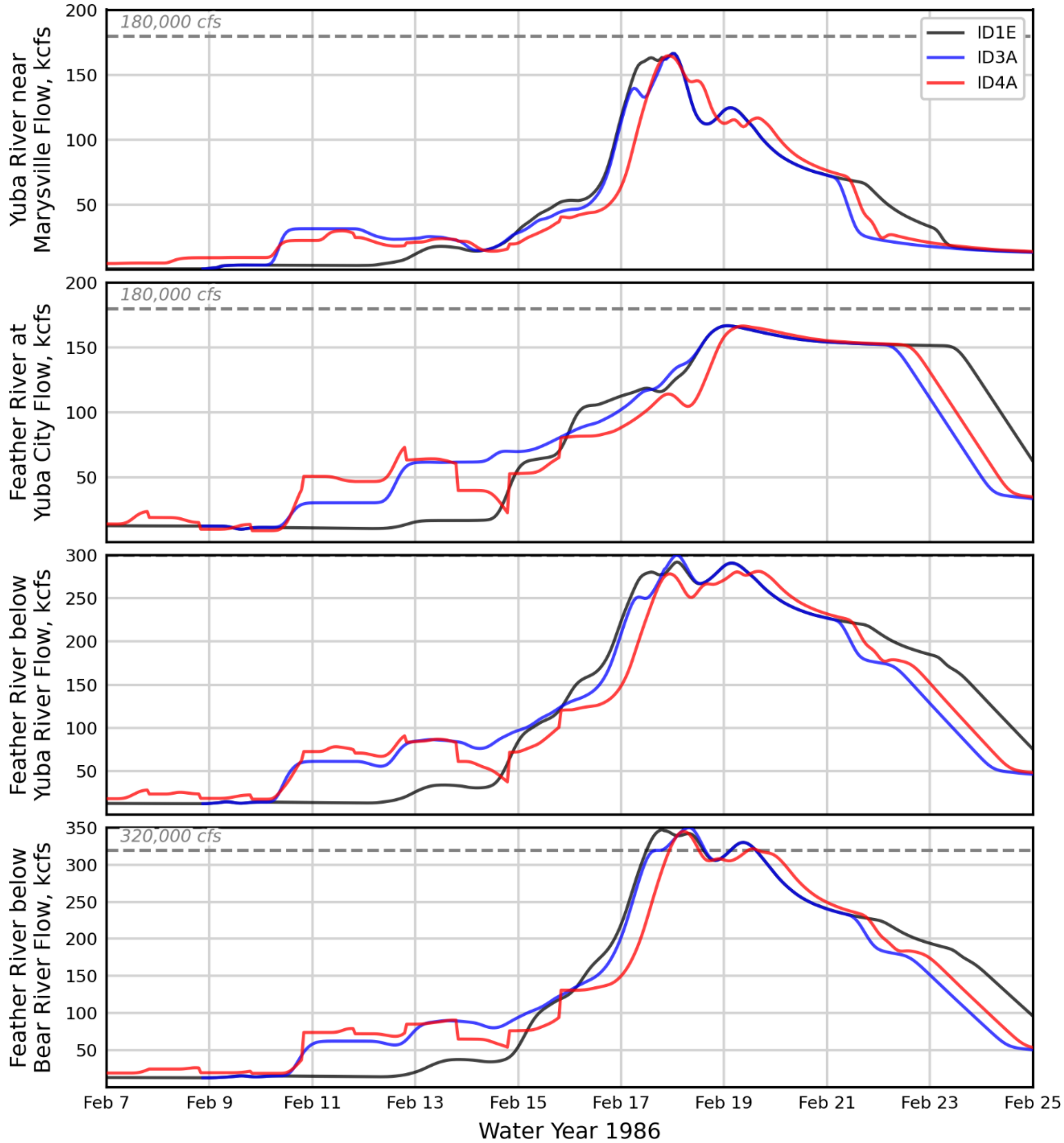
1986 112% Scaling
Oroville



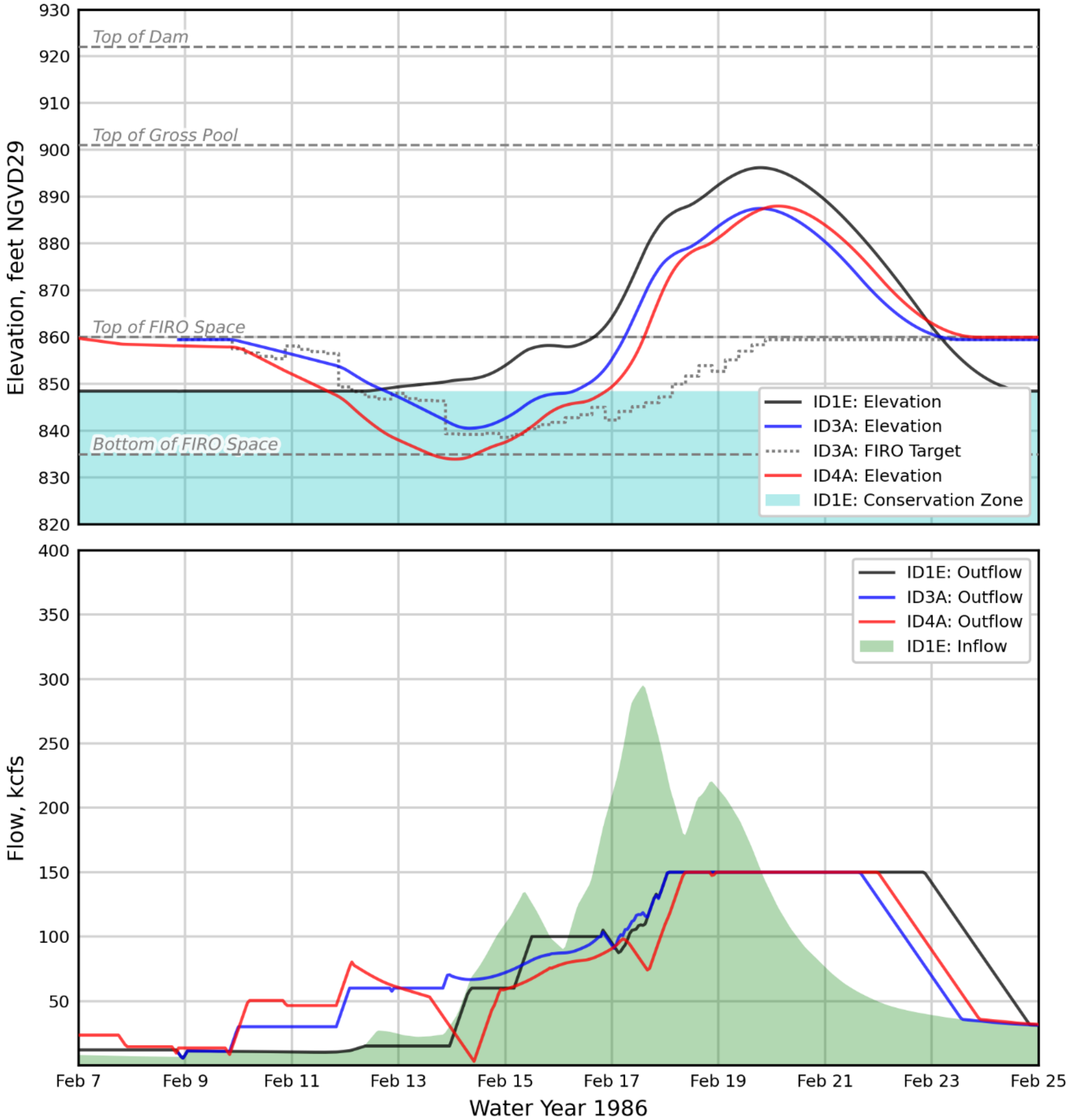
1986 112% Scaling
New Bullards Bar



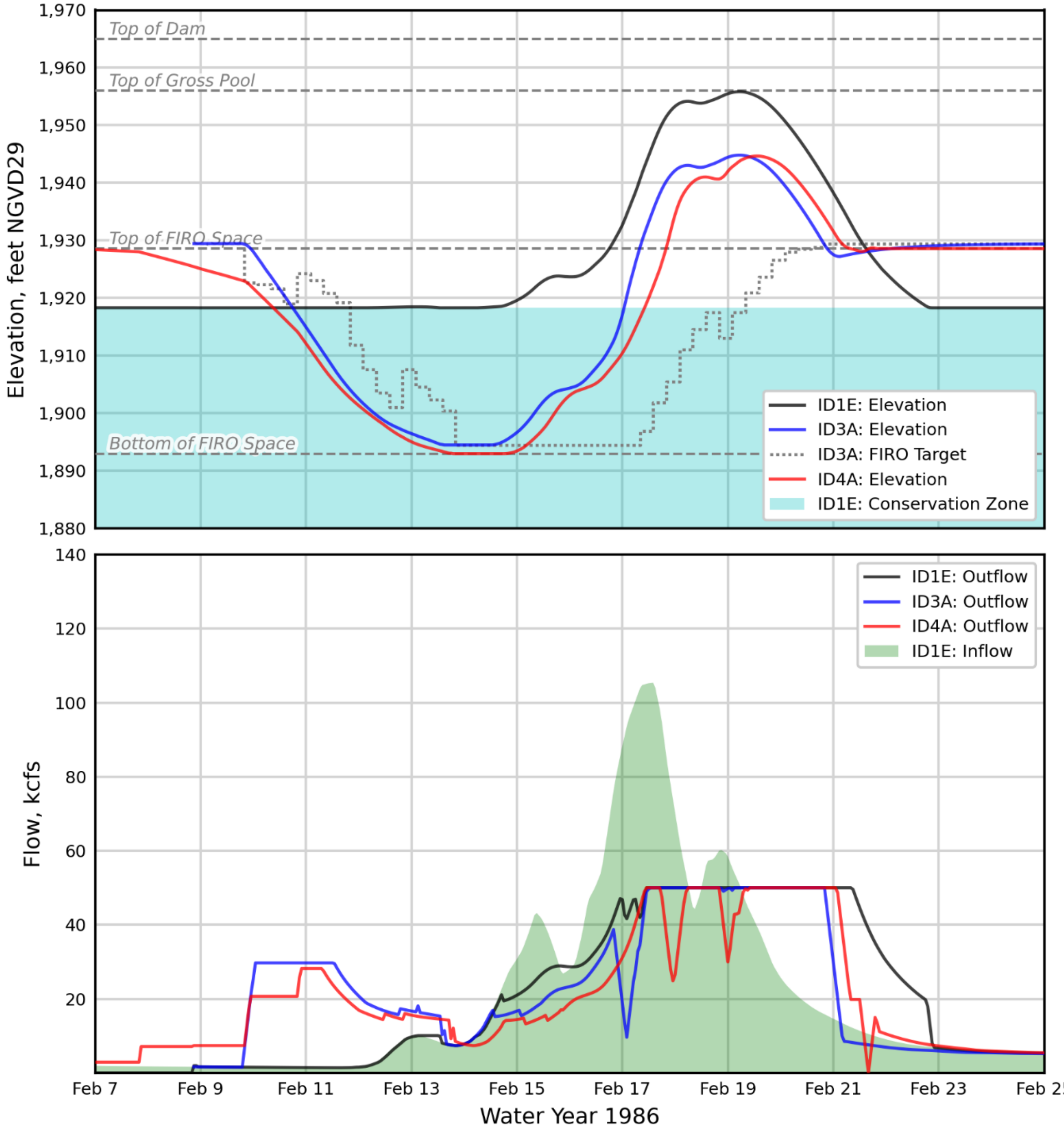
1986 114% Scaling
Downstream Control Points



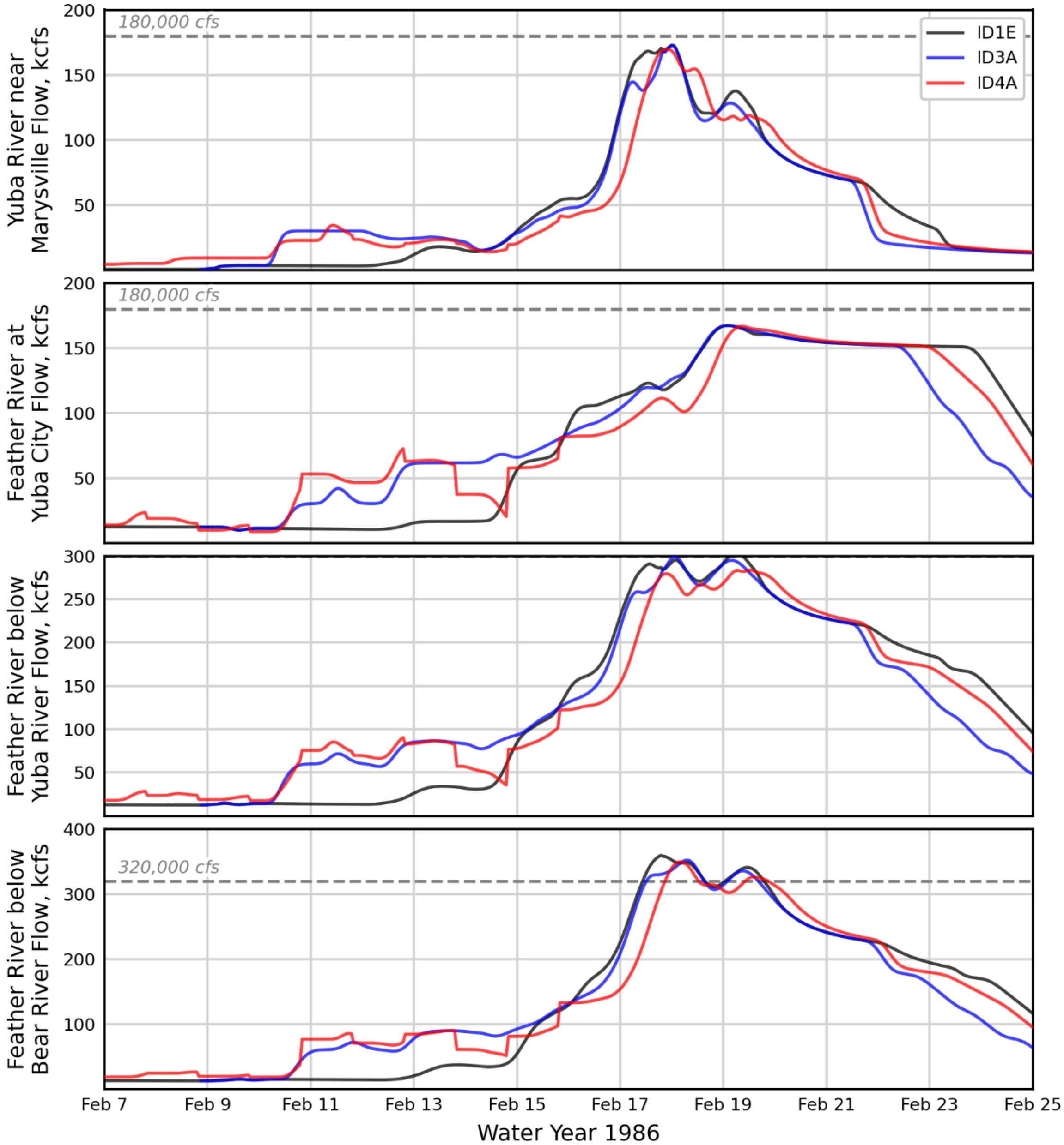
1986 114% Scaling
Oroville



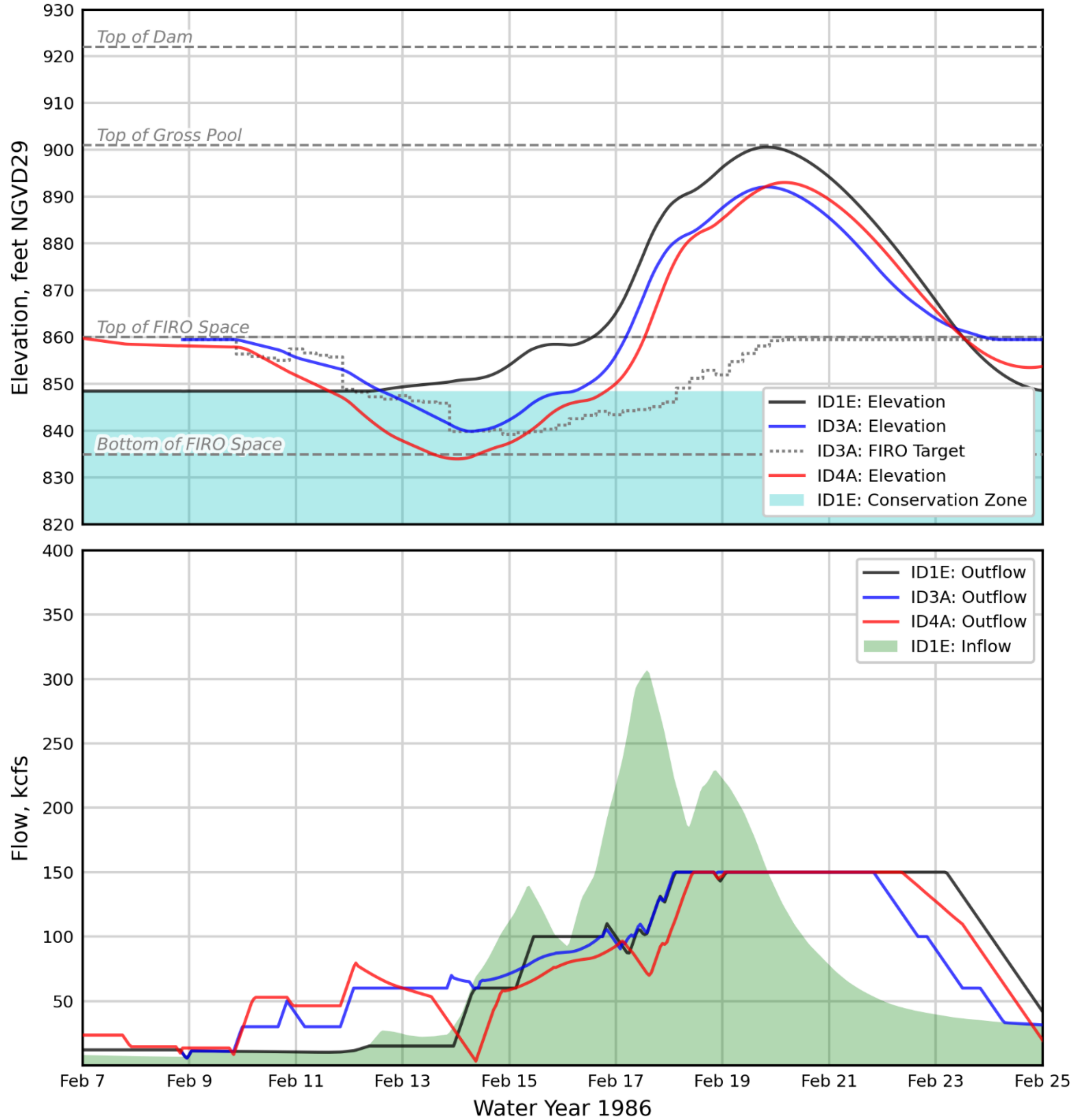
1986 114% Scaling
New Bullards Bar



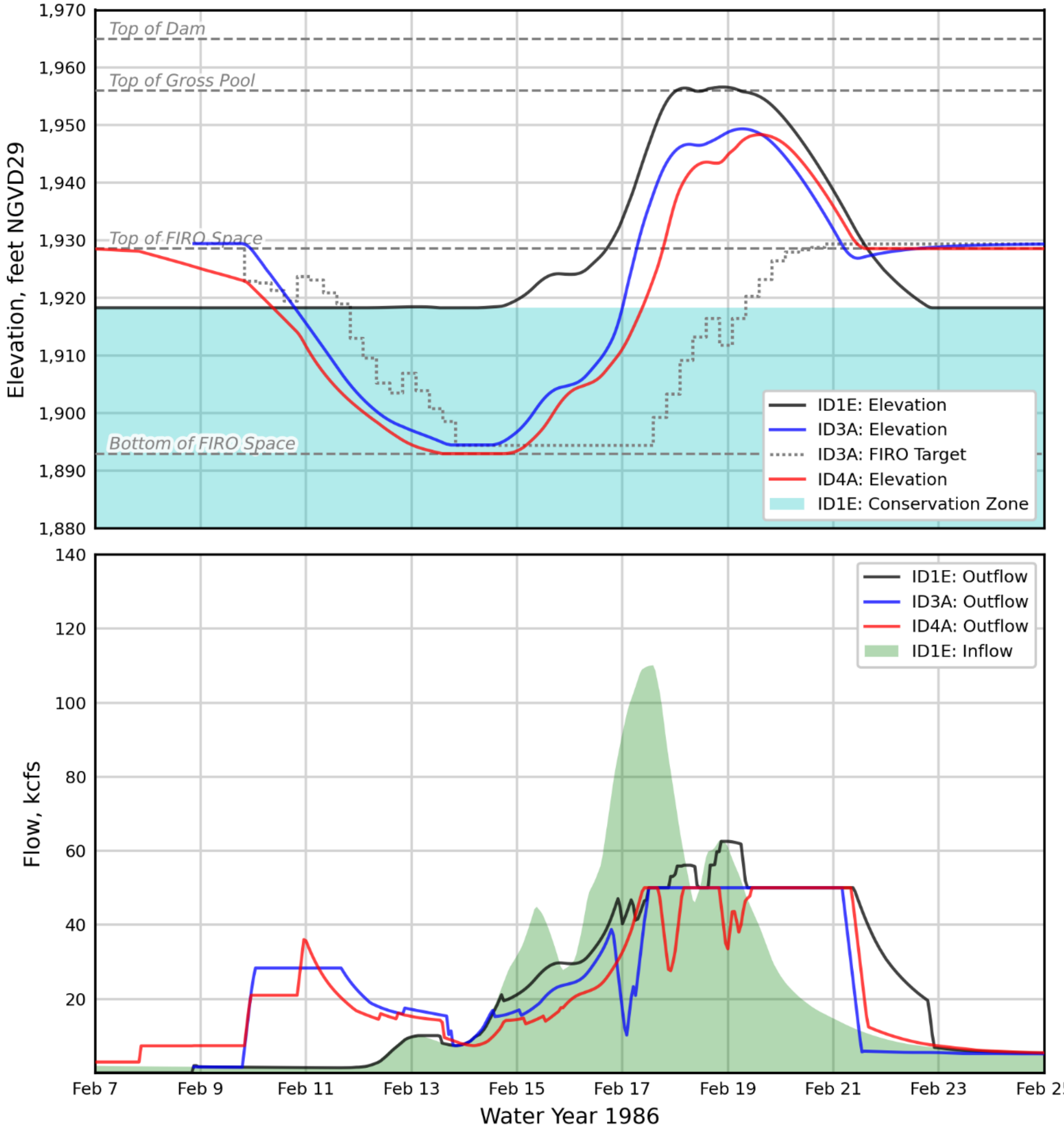
1986 116% Scaling
Downstream Control Points



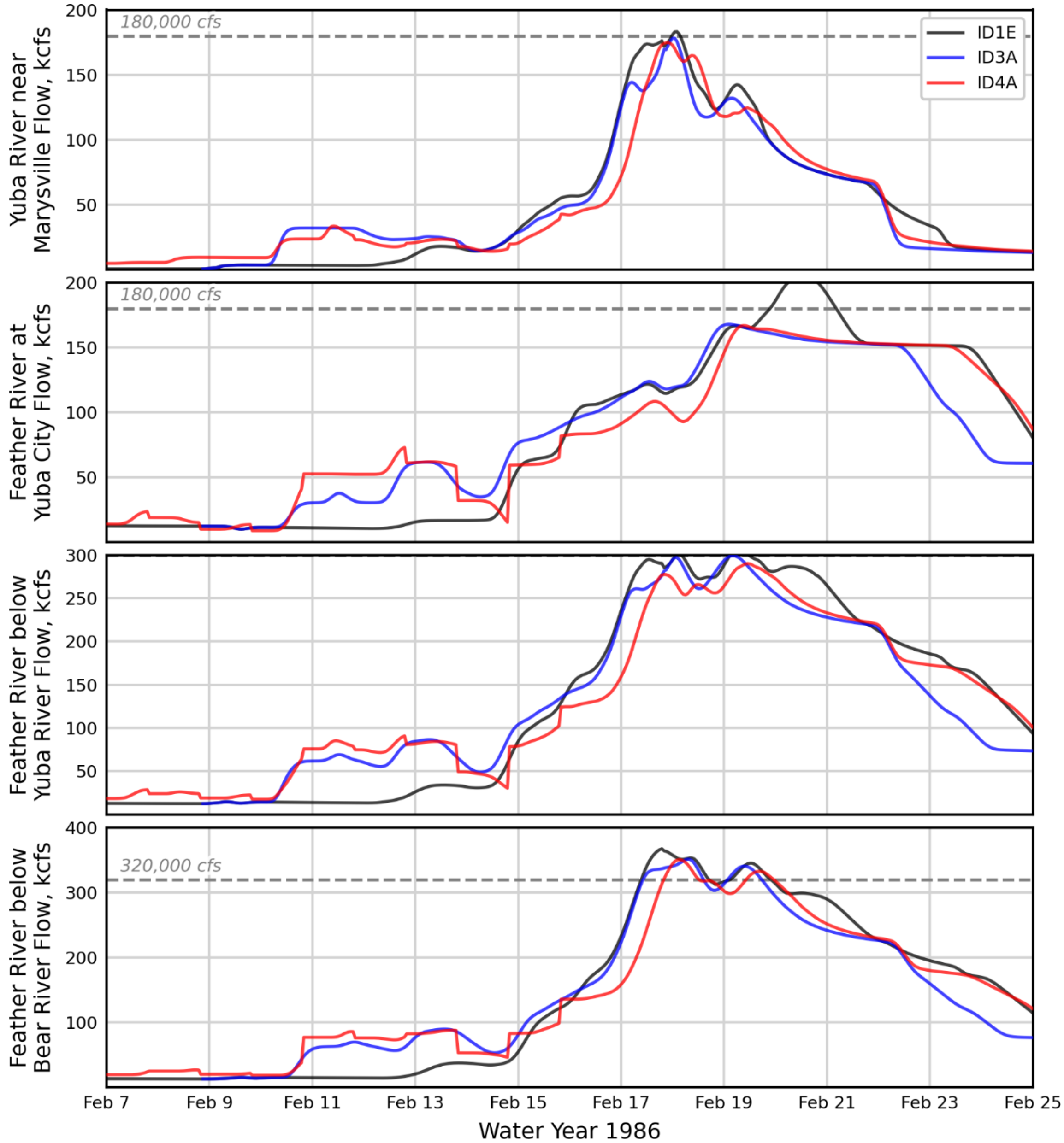
1986 116% Scaling
Oroville



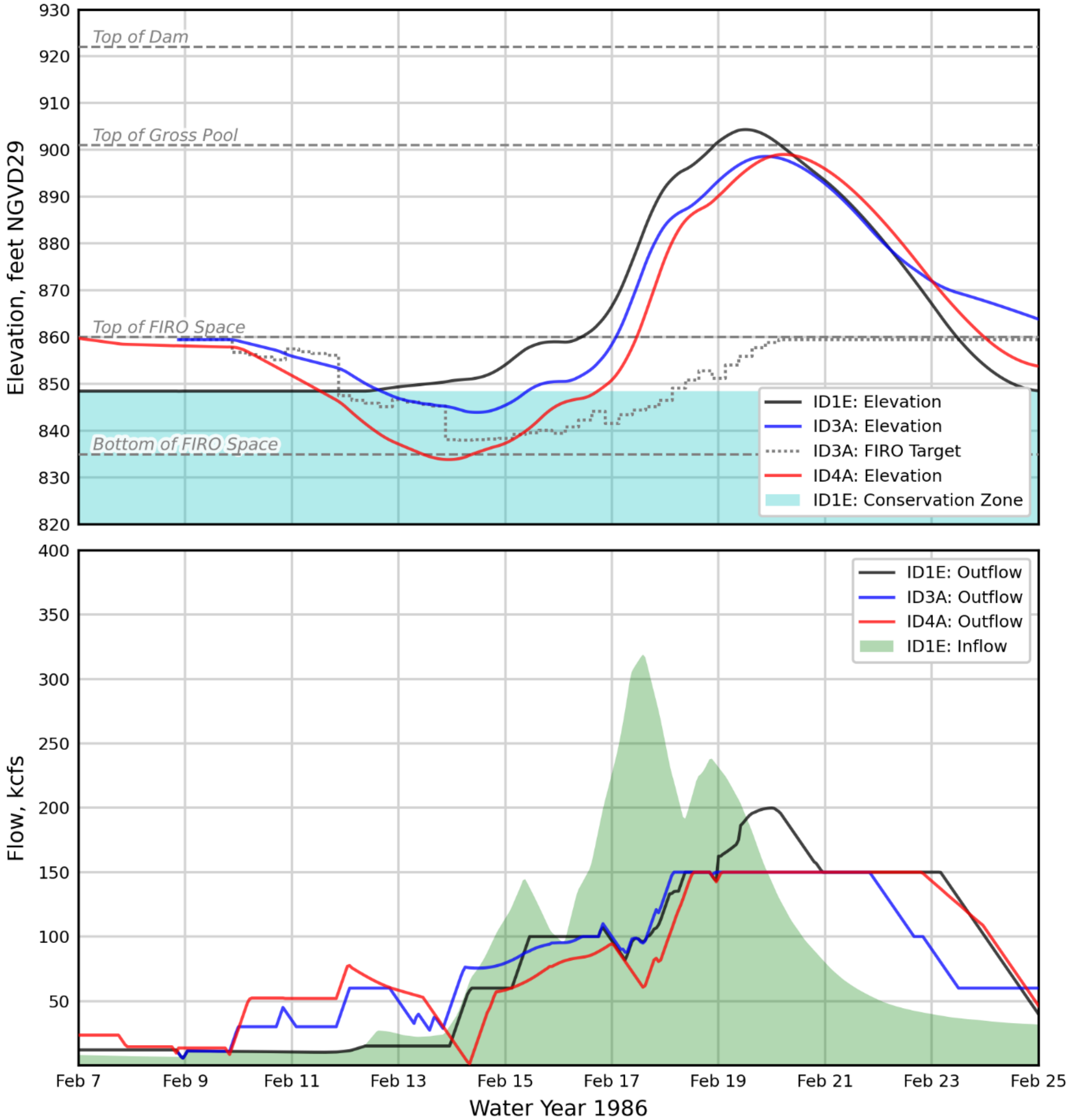
1986 116% Scaling
New Bullards Bar



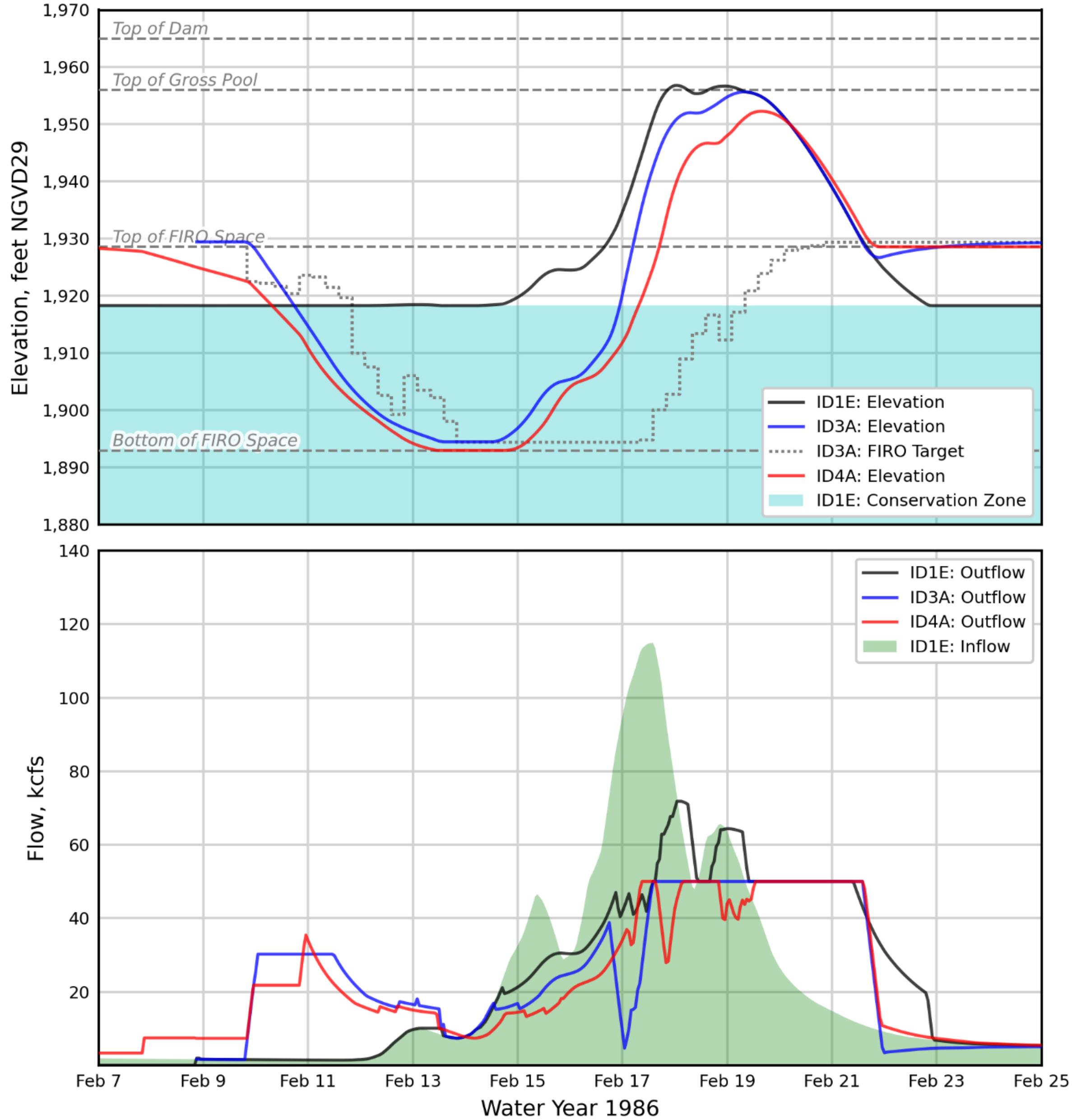
1986 118% Scaling
Downstream Control Points



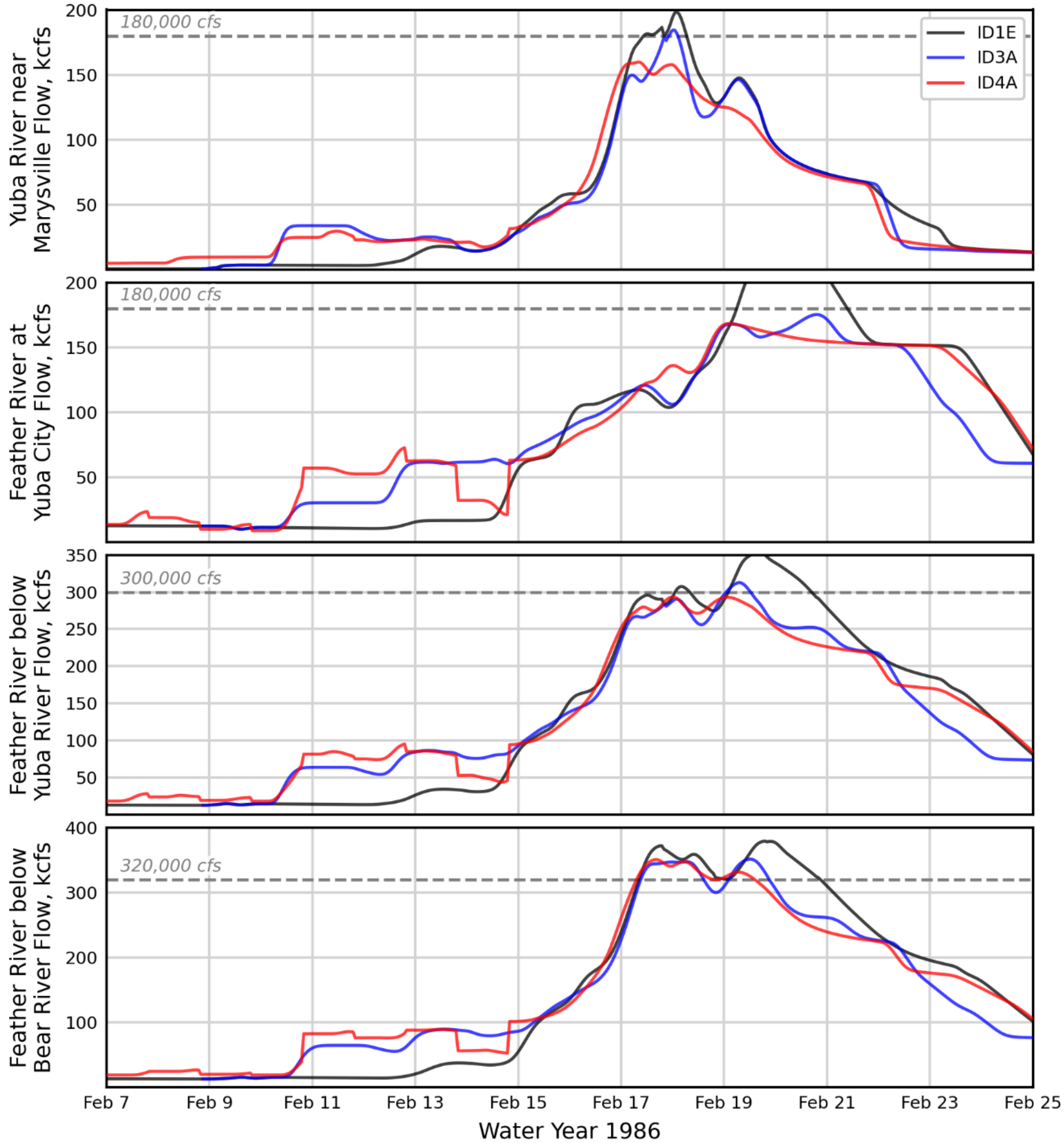
1986 118% Scaling
Oroville



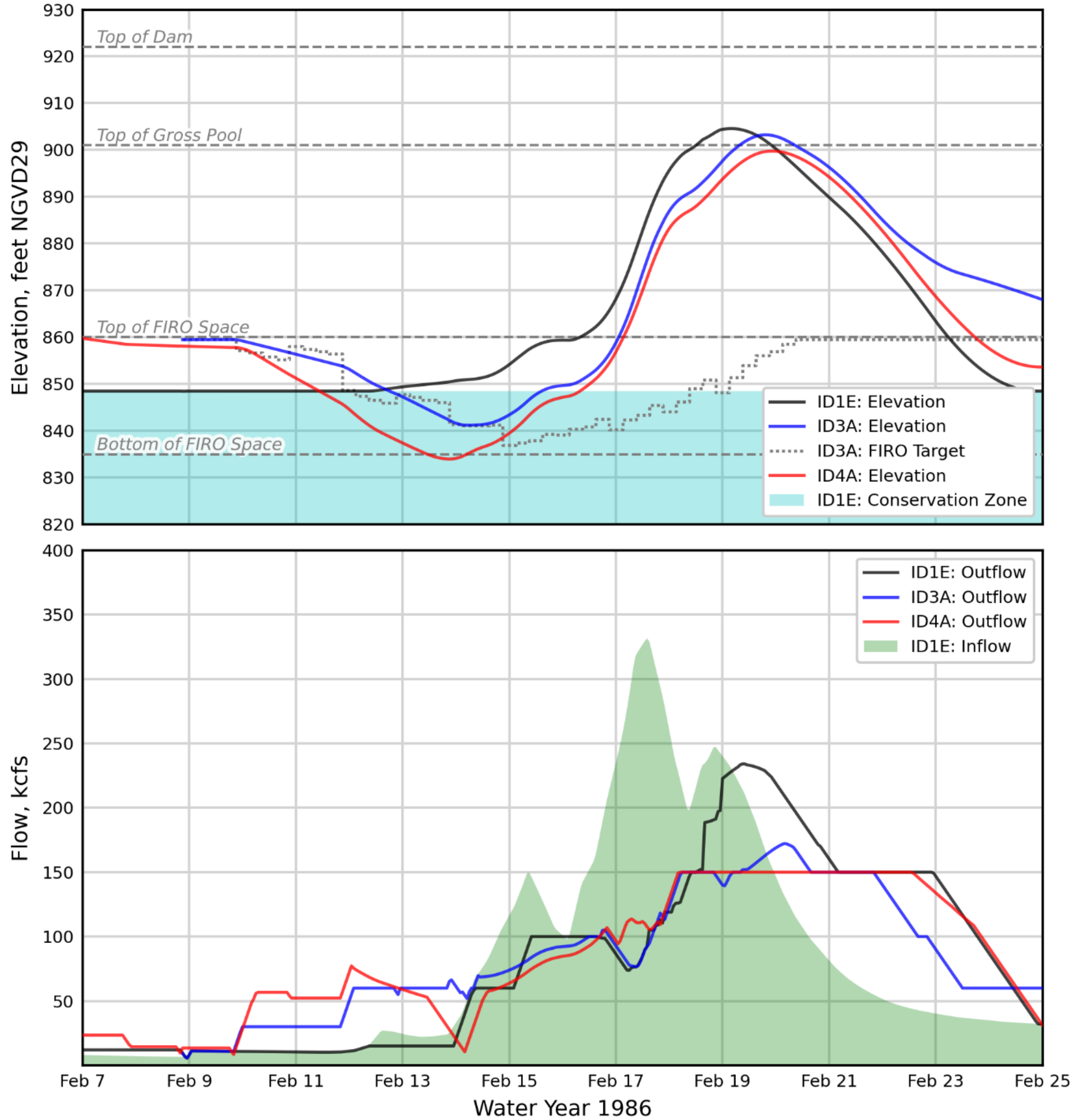
1986 118% Scaling
New Bullards Bar



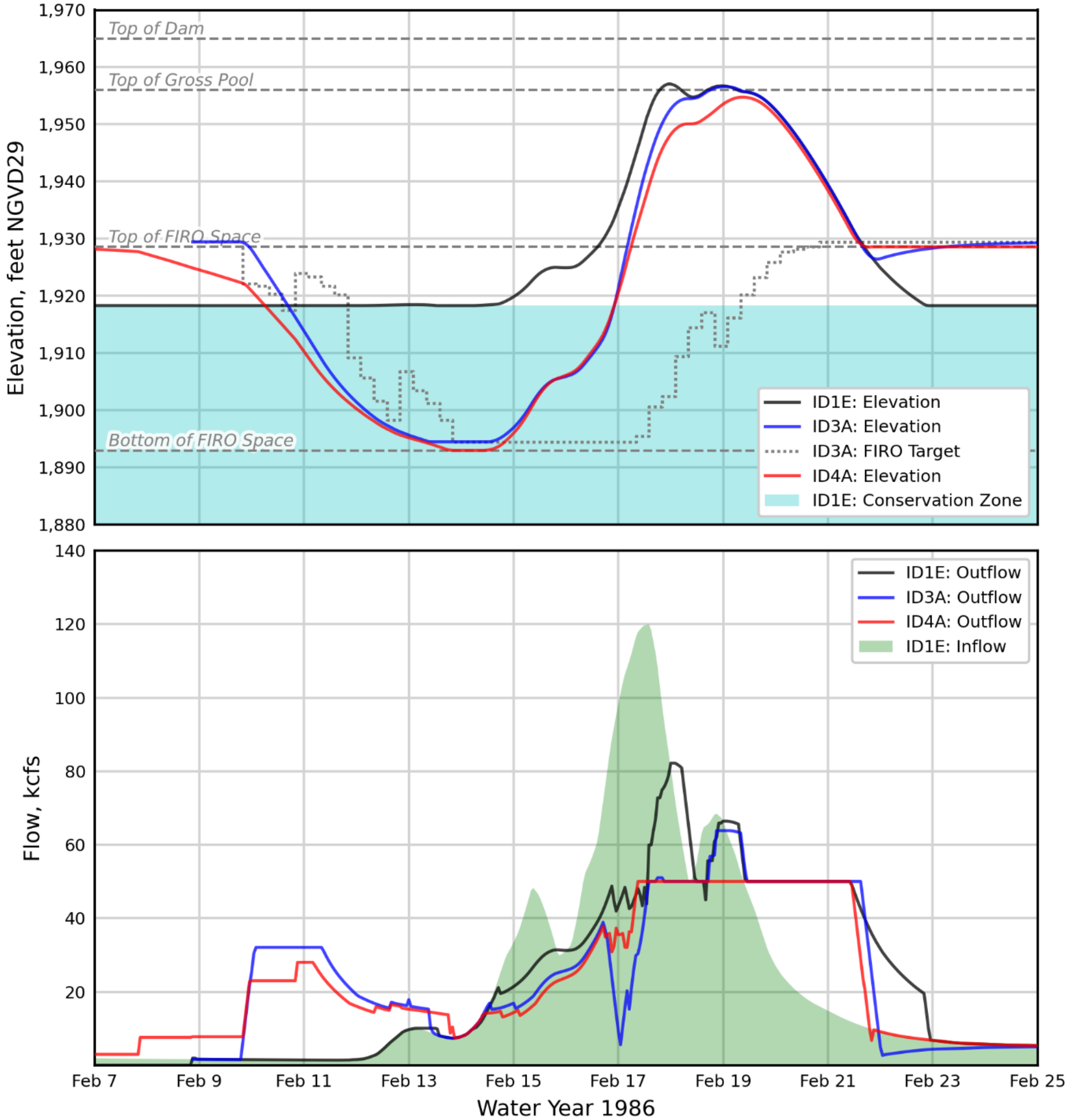
1986 120% Scaling
Downstream Control Points



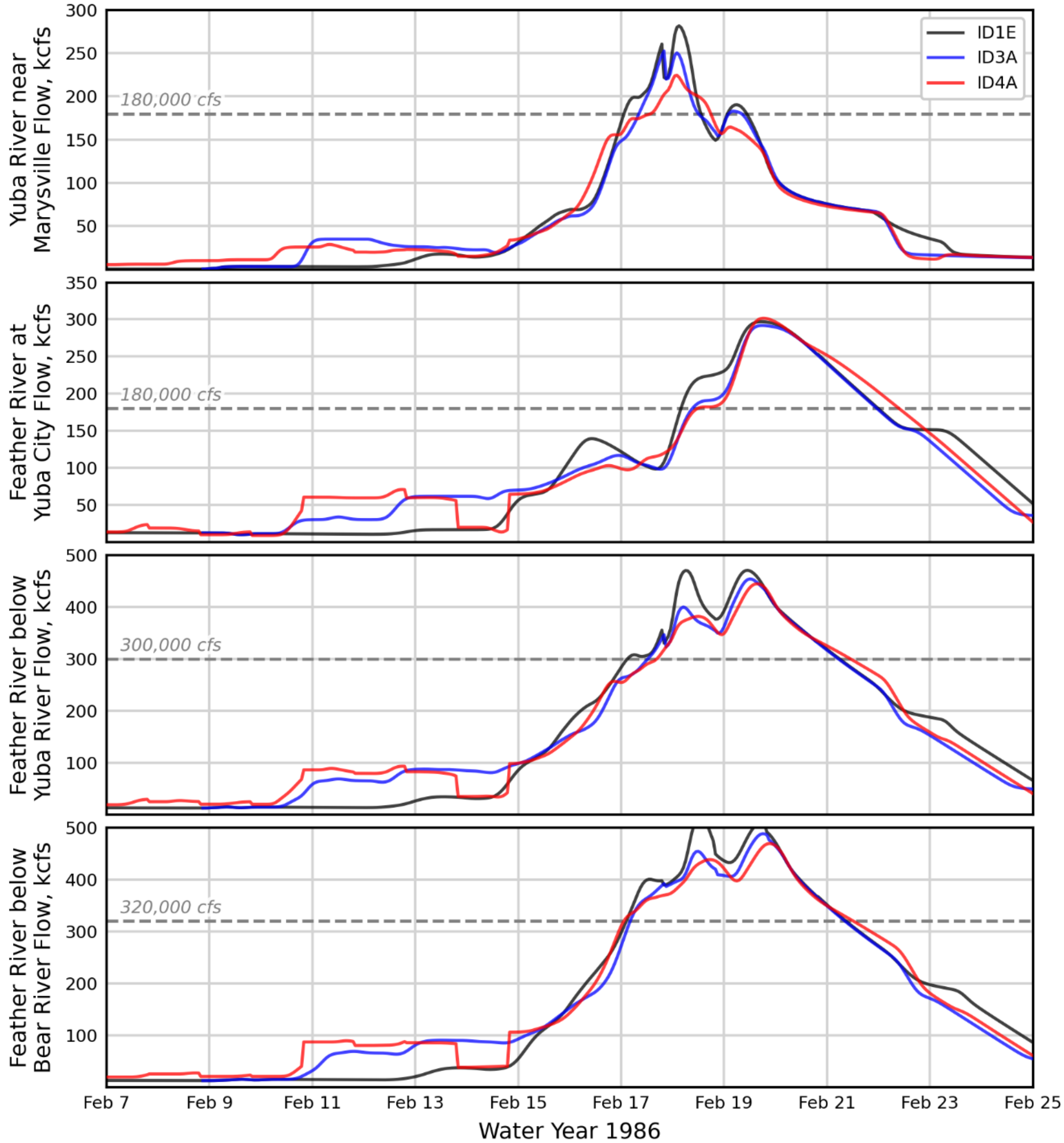
1986 120% Scaling
Oroville



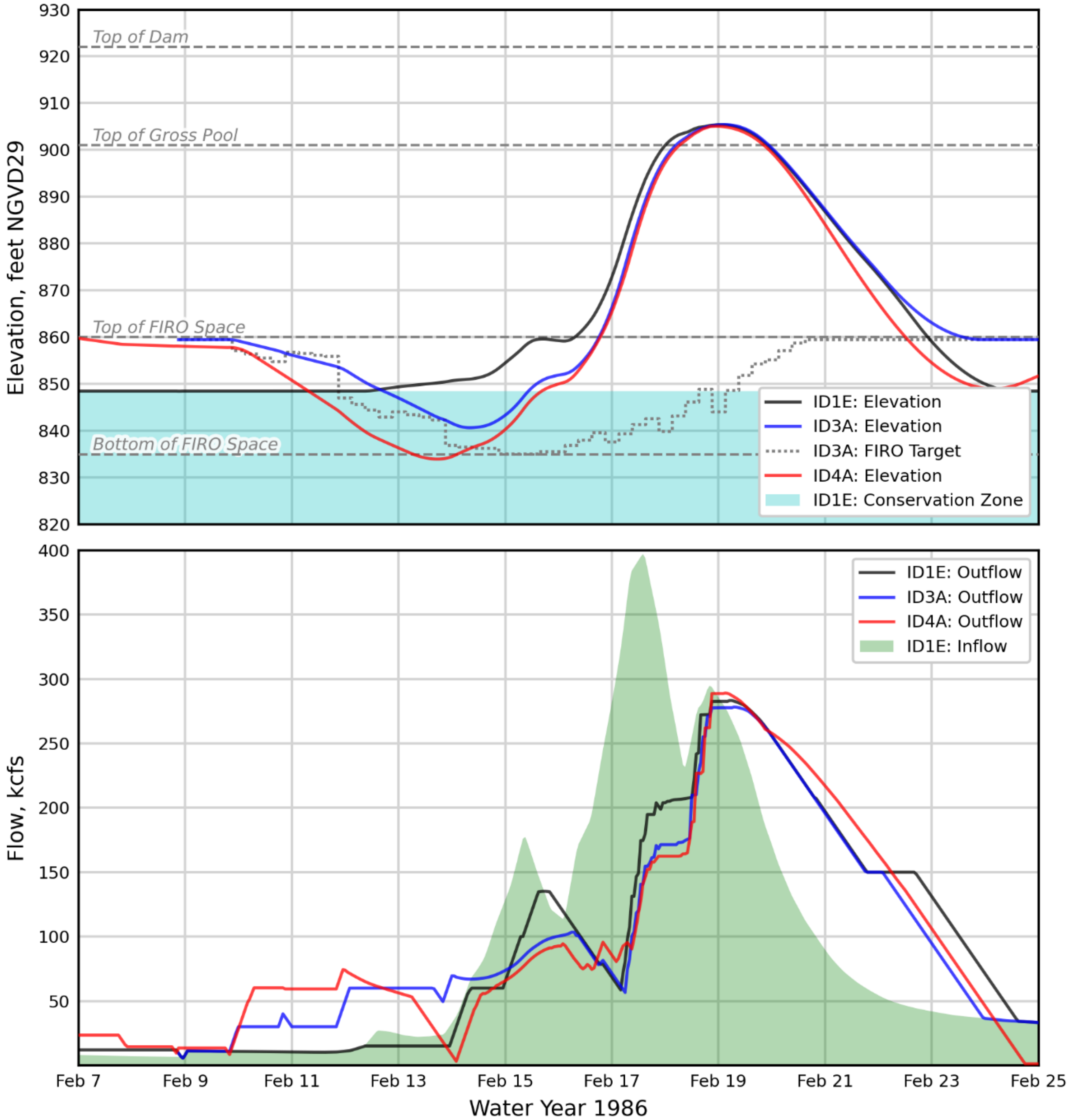
1986 120% Scaling
New Bullards Bar



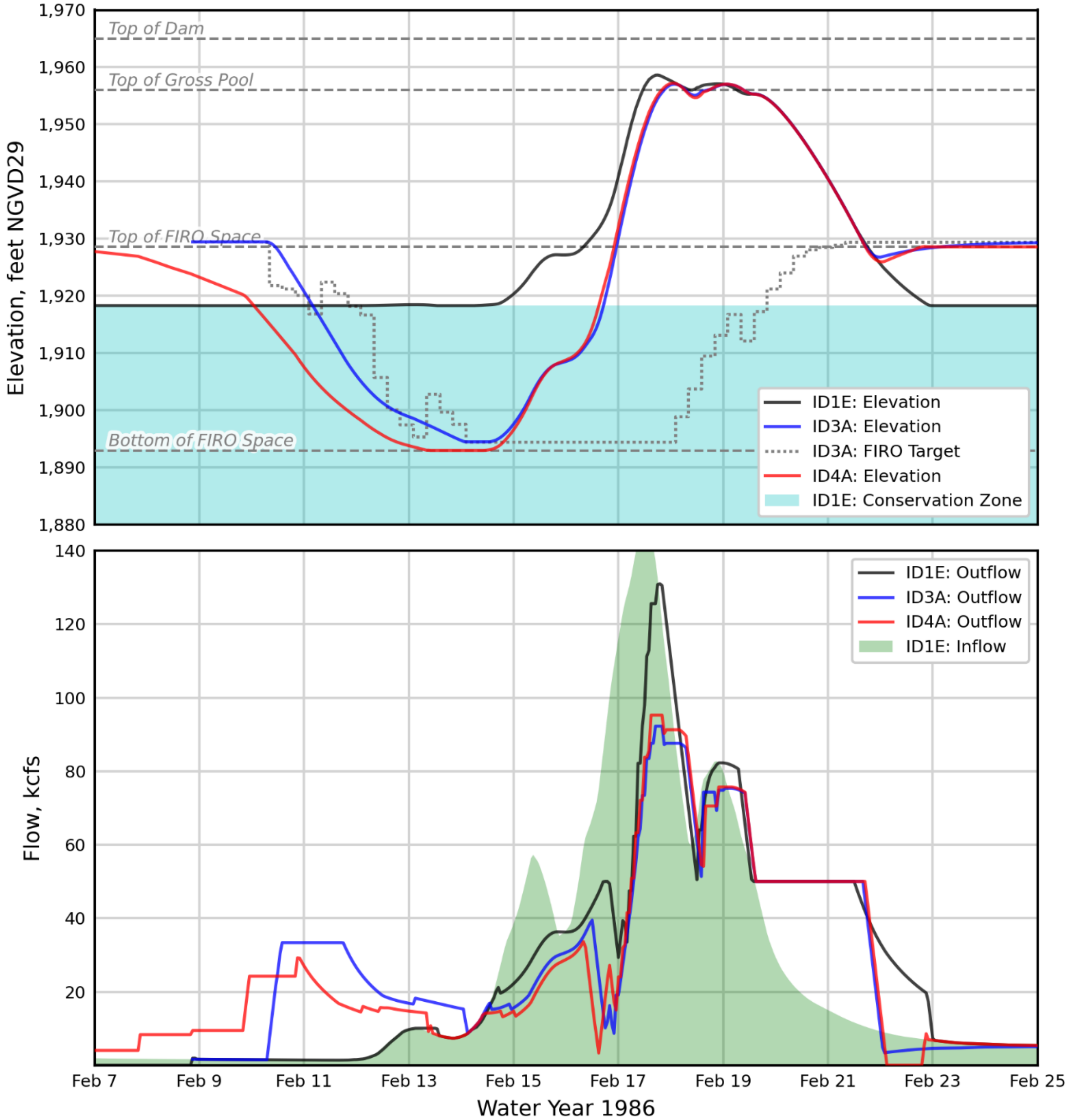
1986 130% Scaling
Downstream Control Points



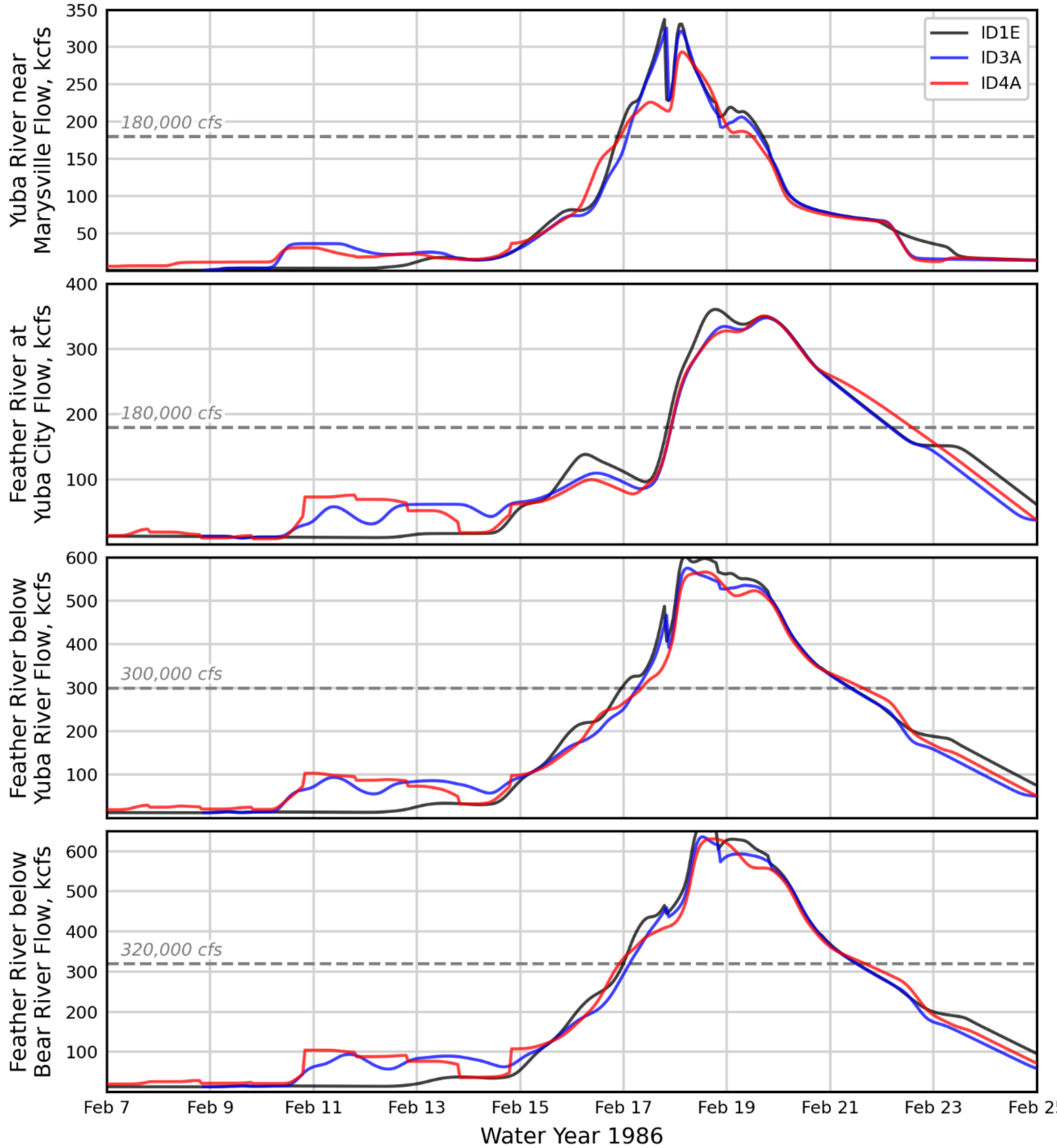
1986 130% Scaling
Oroville



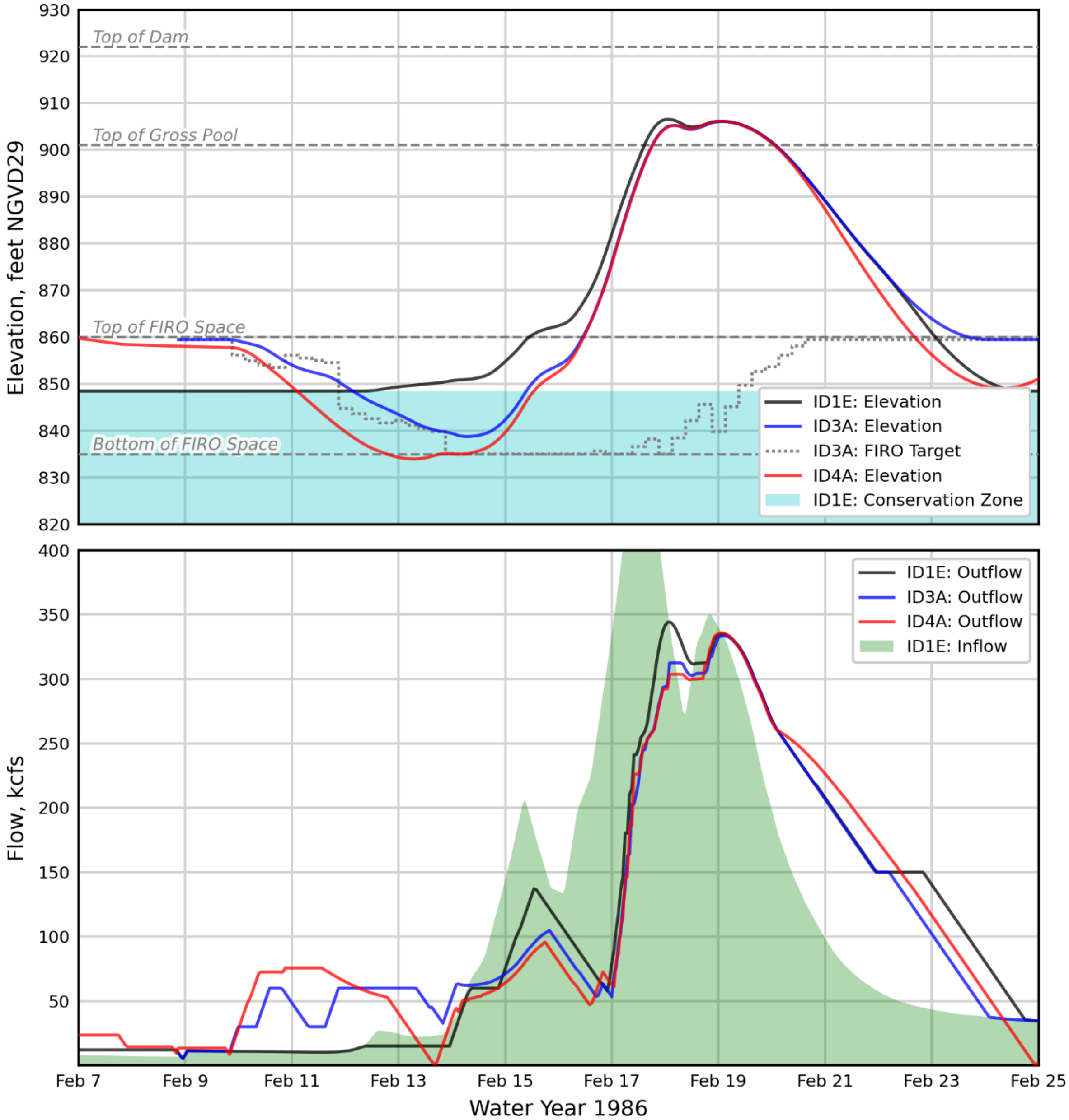
1986 130% Scaling
New Bullards Bar



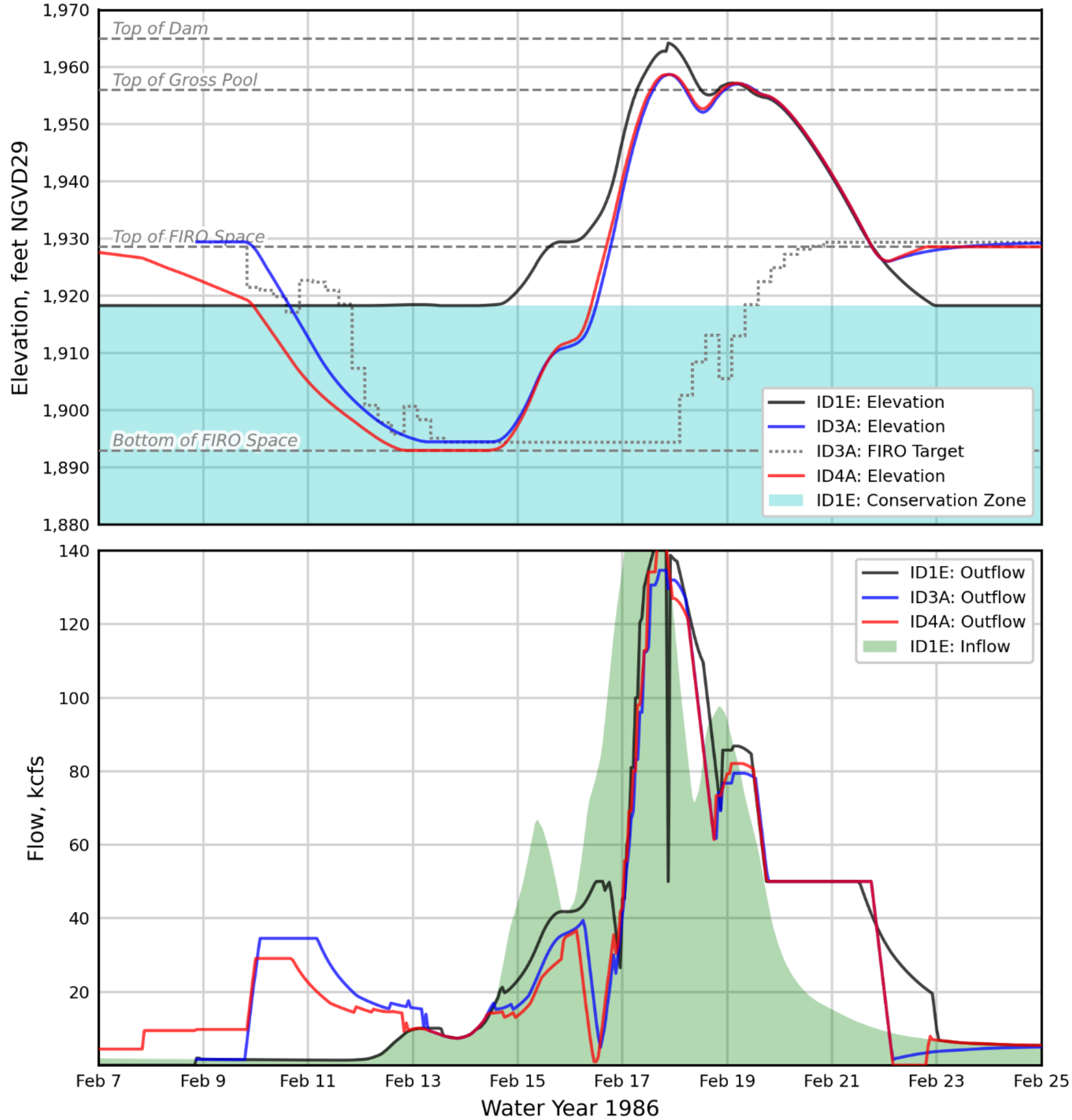
1986 140% Scaling
Downstream Control Points



1986 140% Scaling
Oroville



1986 140% Scaling
New Bullards Bar



A.5 Scalings of the 1997 Flood Event

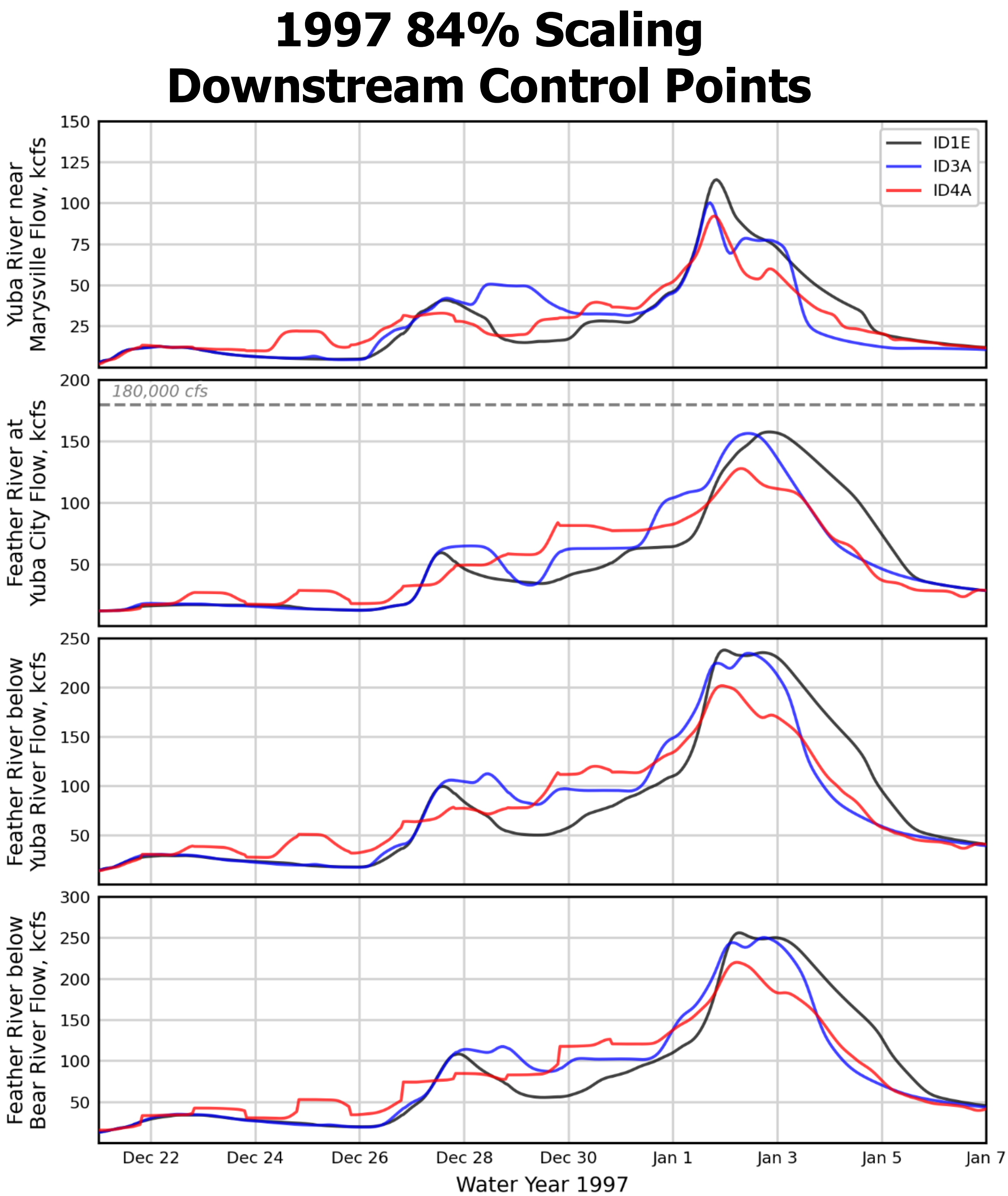
The scaling process is described in Section 5.3.2.2

Scale factors applied: 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 120, 130

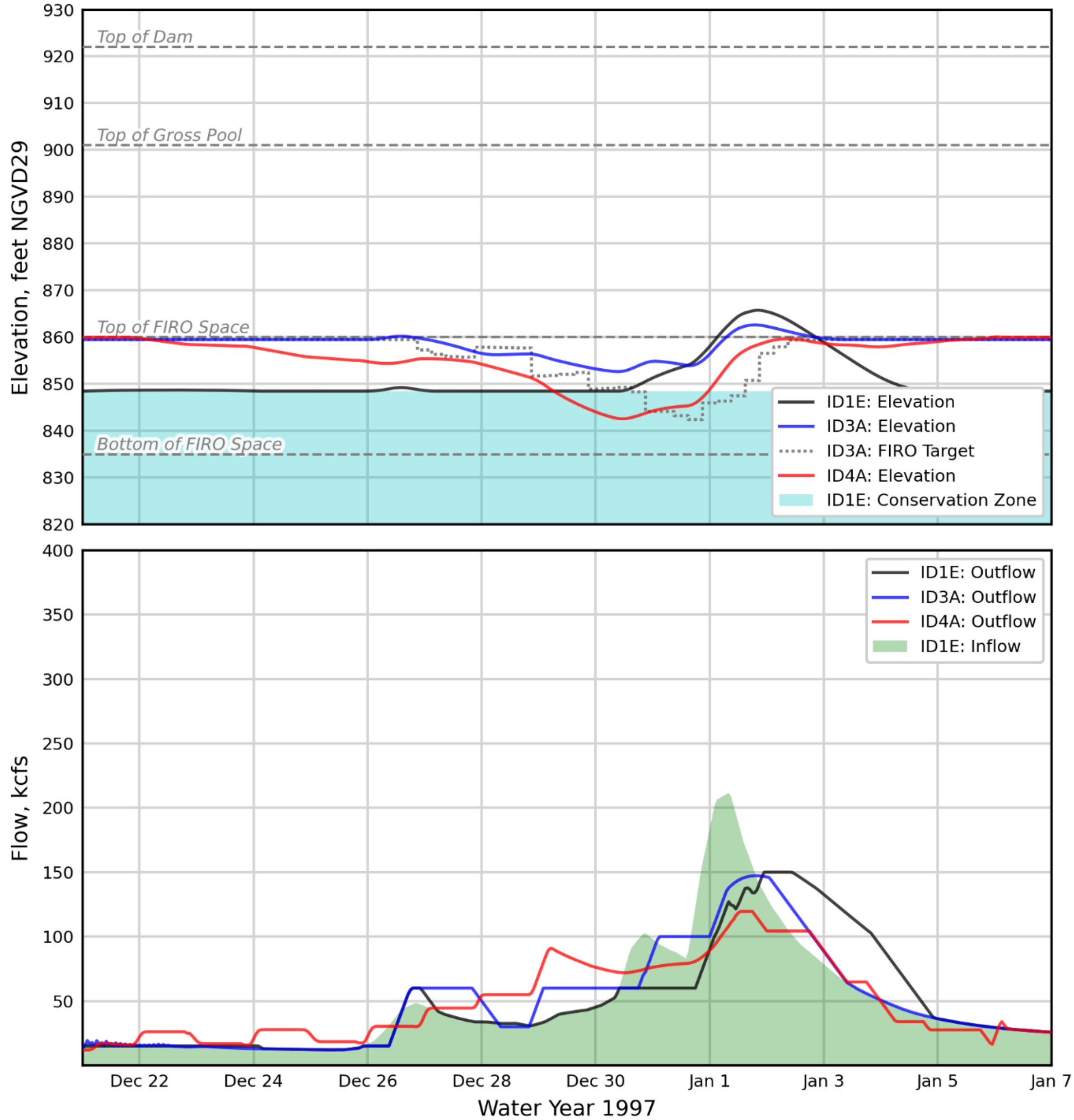
Each scaling contains:

- A plot of the four downstream control points
- A plot of Oroville elevation (top) and releases (bottom) with inflow as the background

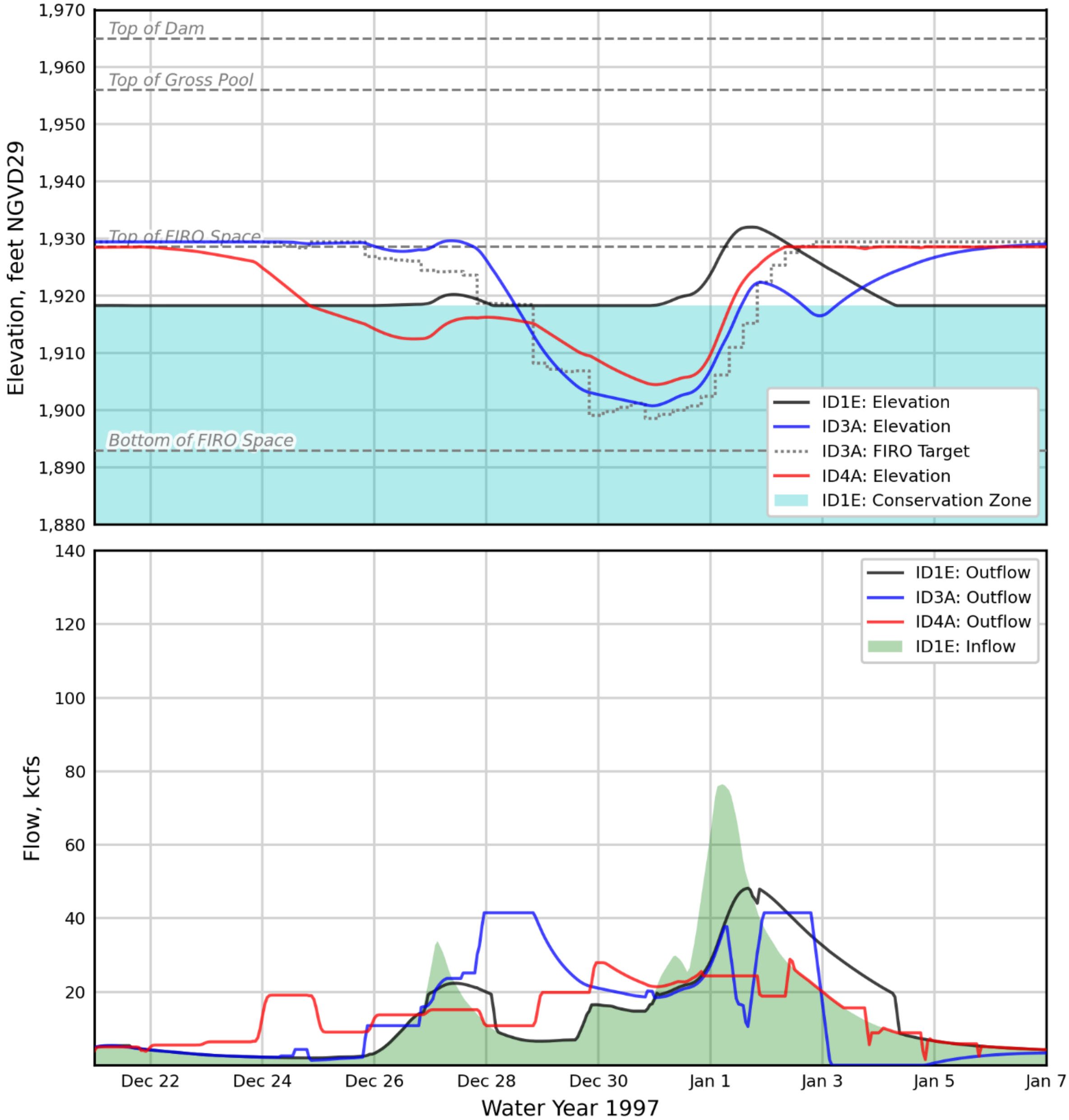
Alternatives (ID1E, ID3A, and ID4A) are described in Section 3.



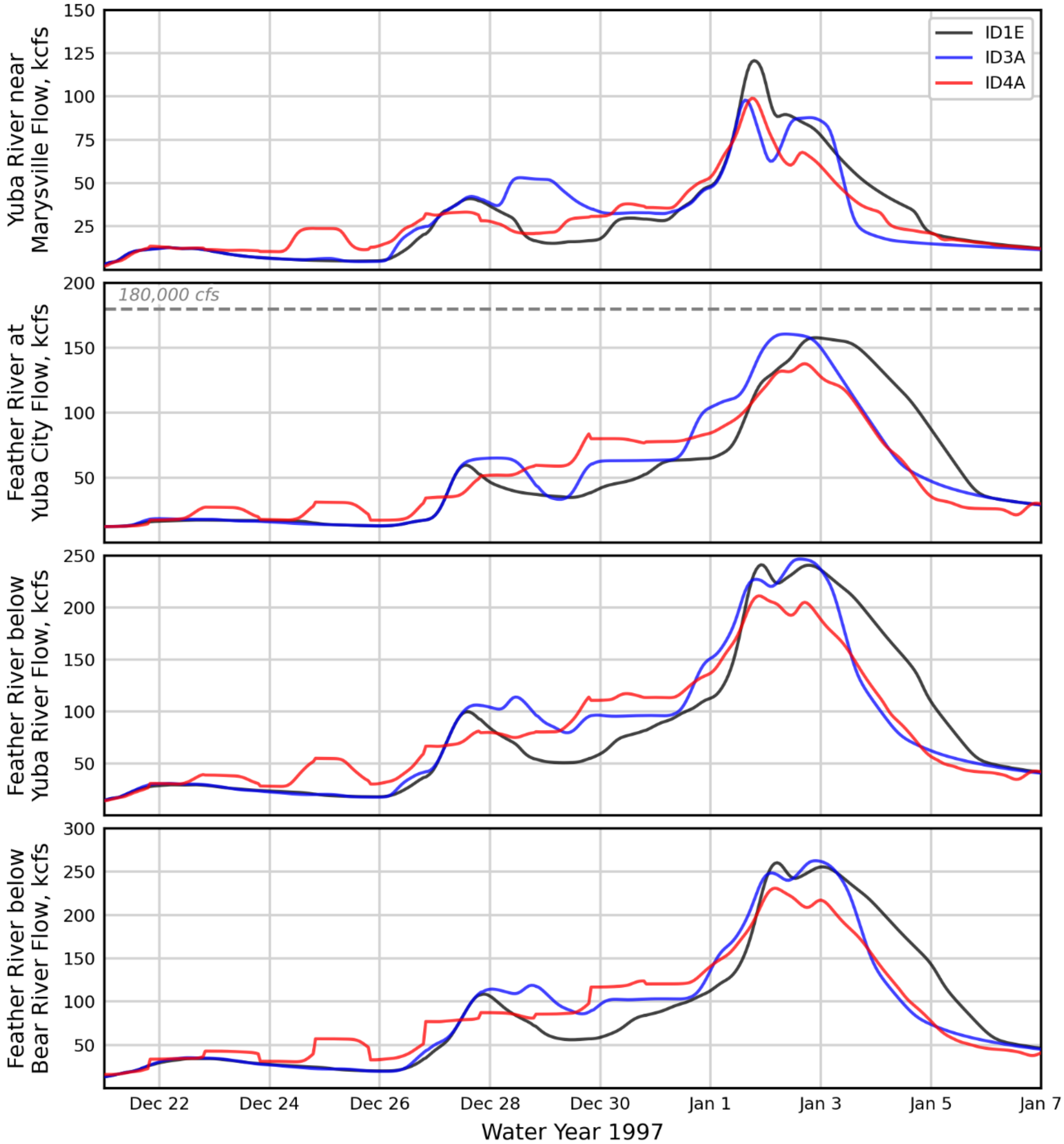
1997 84% Scaling
Oroville



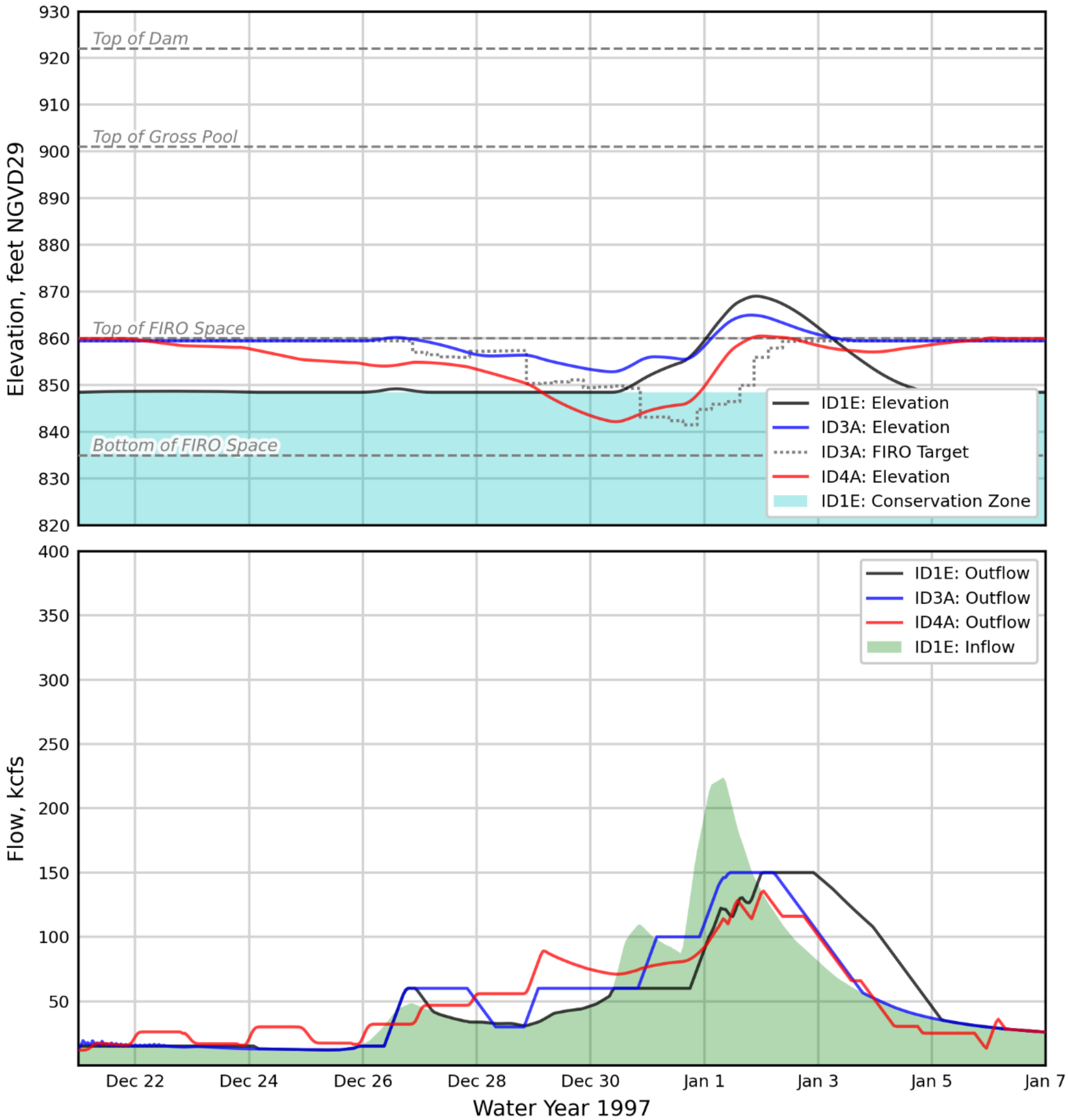
1997 84% Scaling
New Bullards Bar



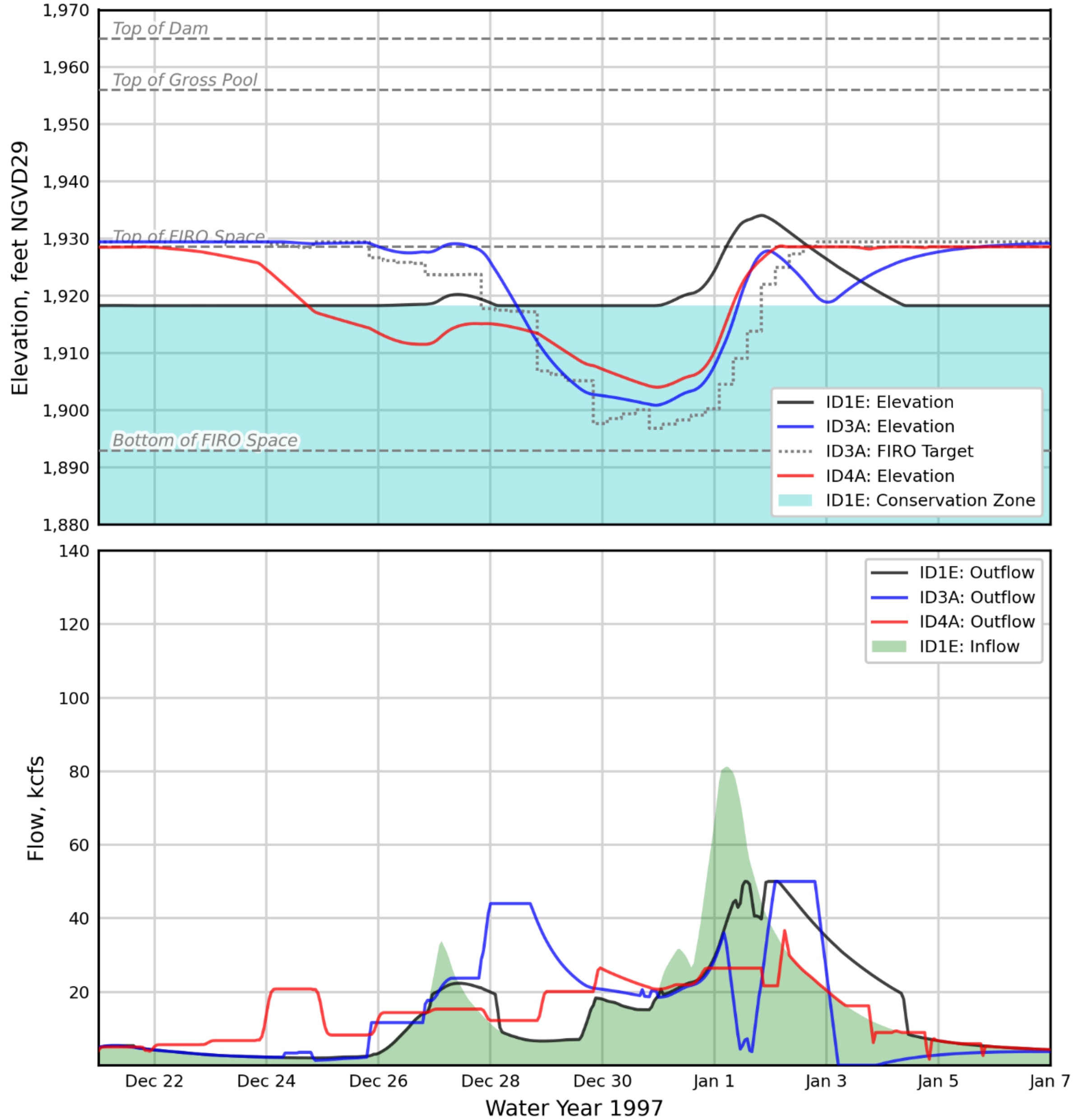
1997 86% Scaling
Downstream Control Points



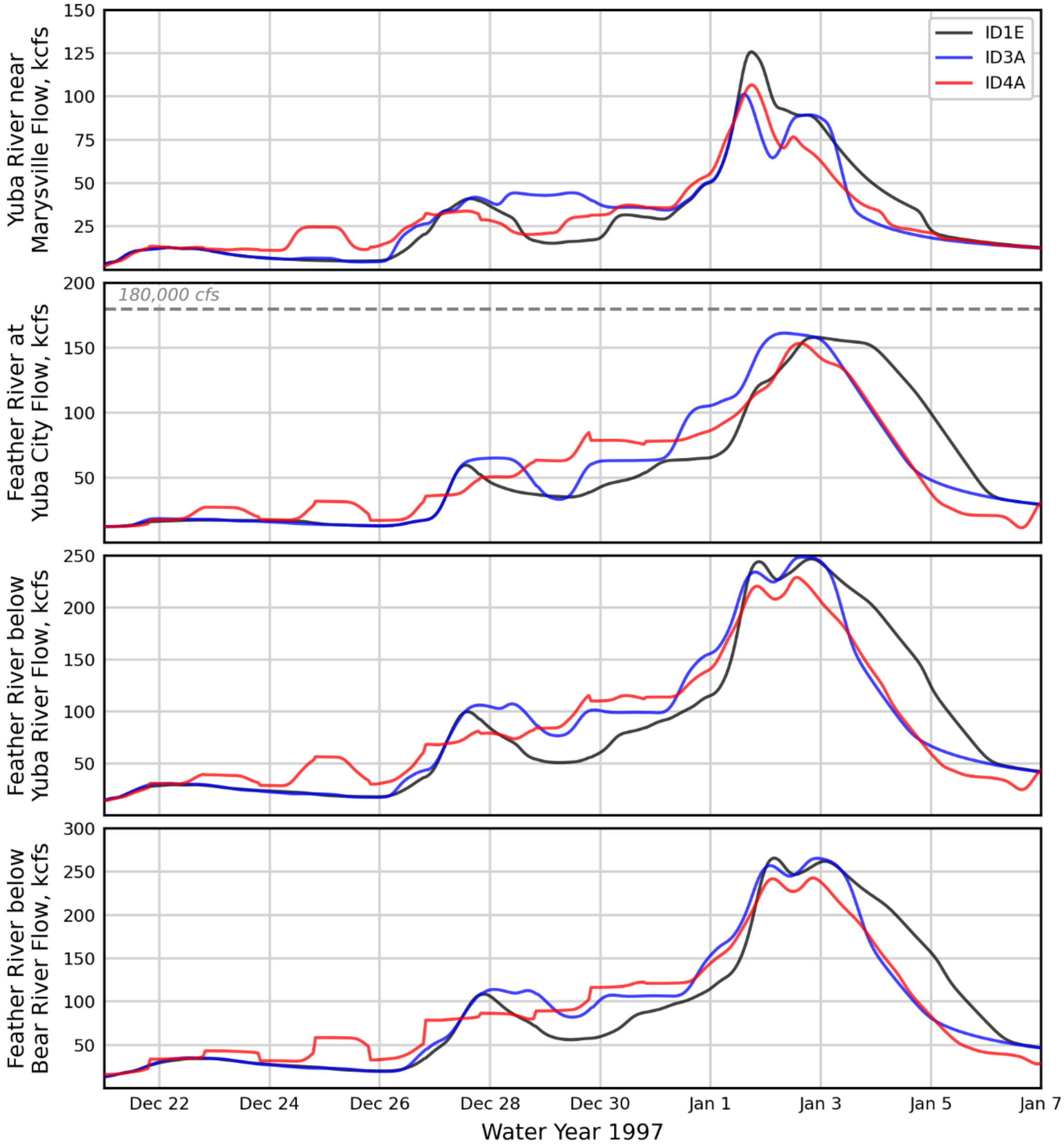
1997 86% Scaling
Oroville



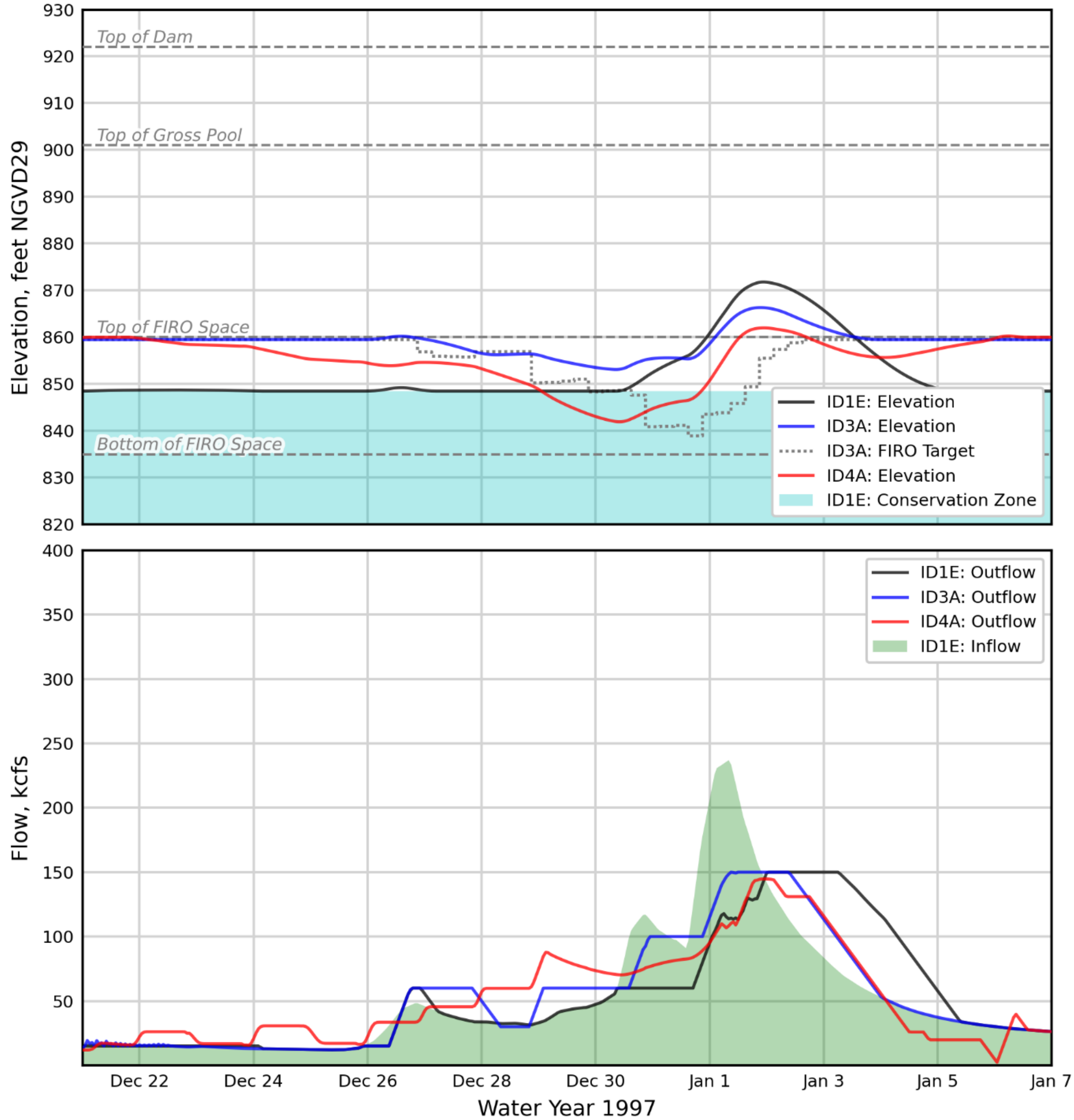
1997 86% Scaling
New Bullards Bar



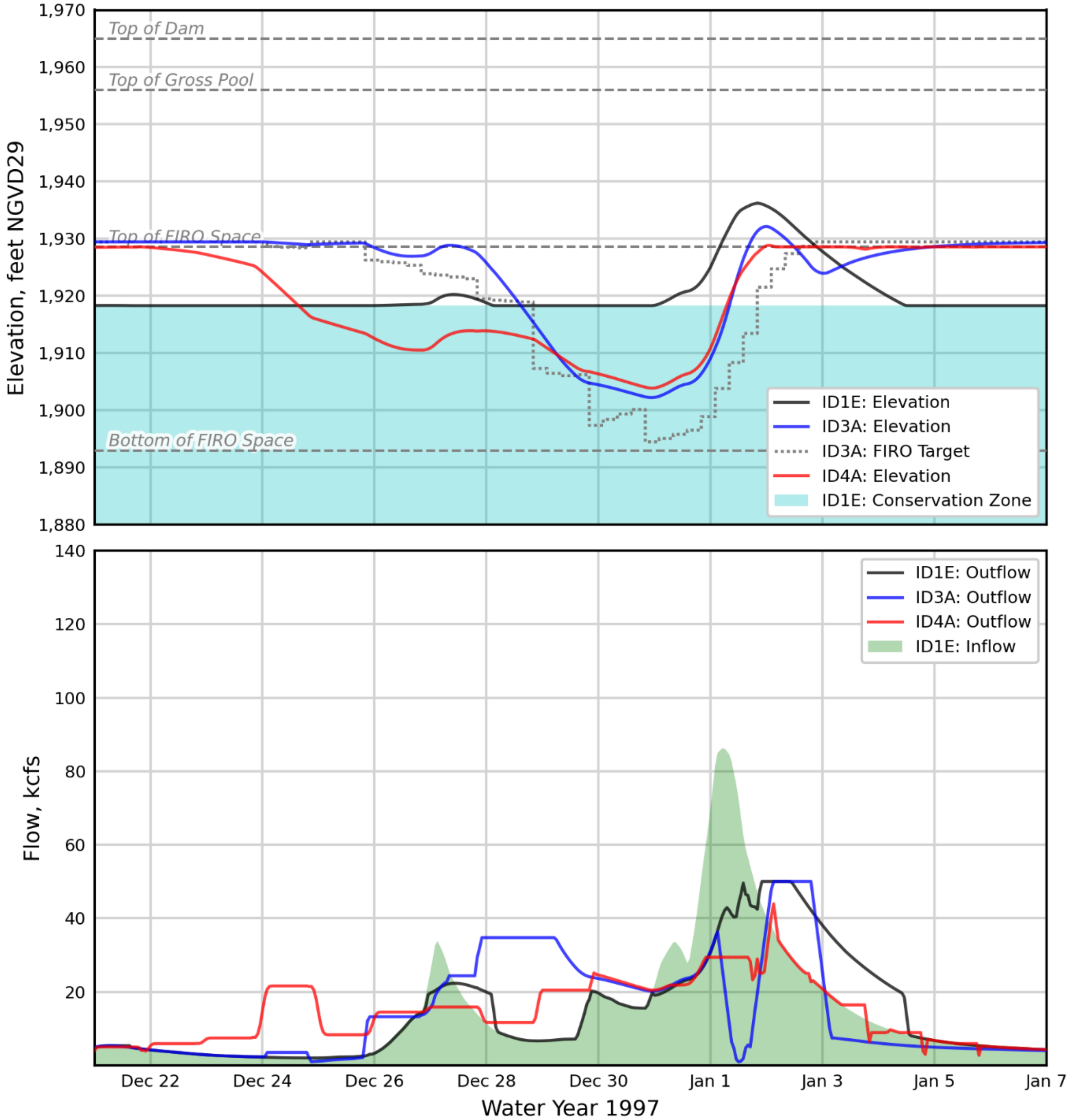
1997 88% Scaling
Downstream Control Points



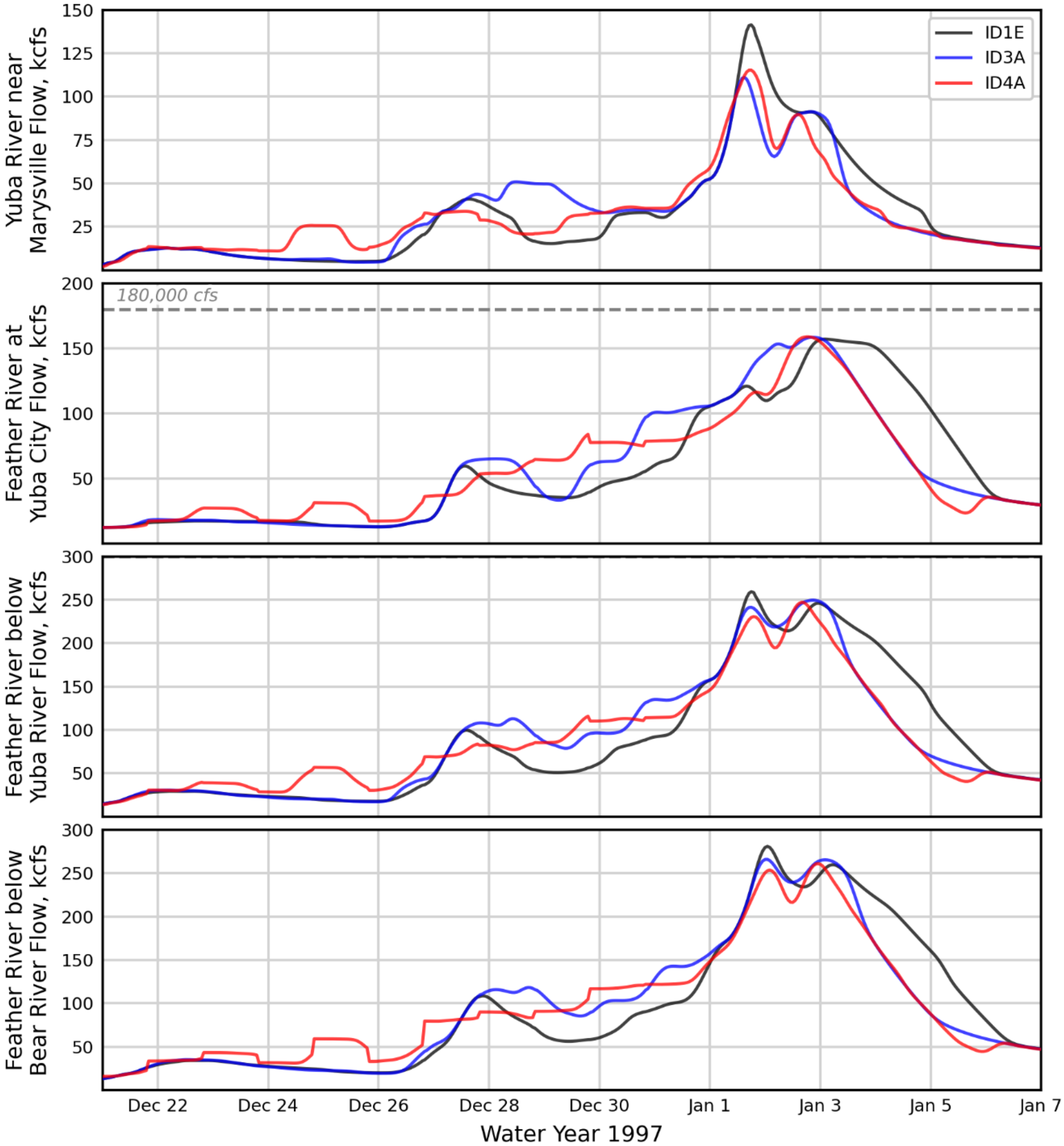
1997 88% Scaling
Oroville



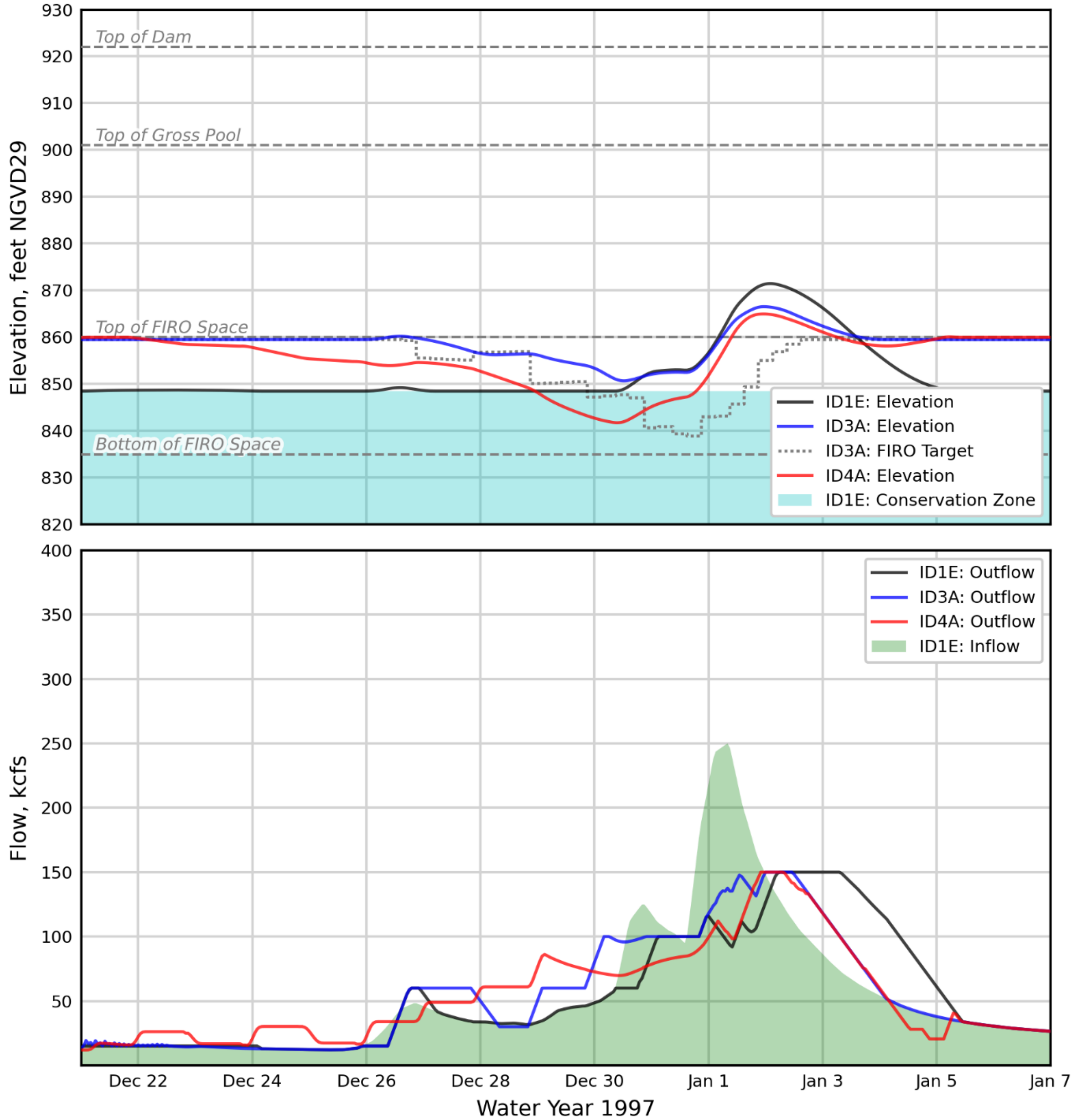
1997 88% Scaling
New Bullards Bar



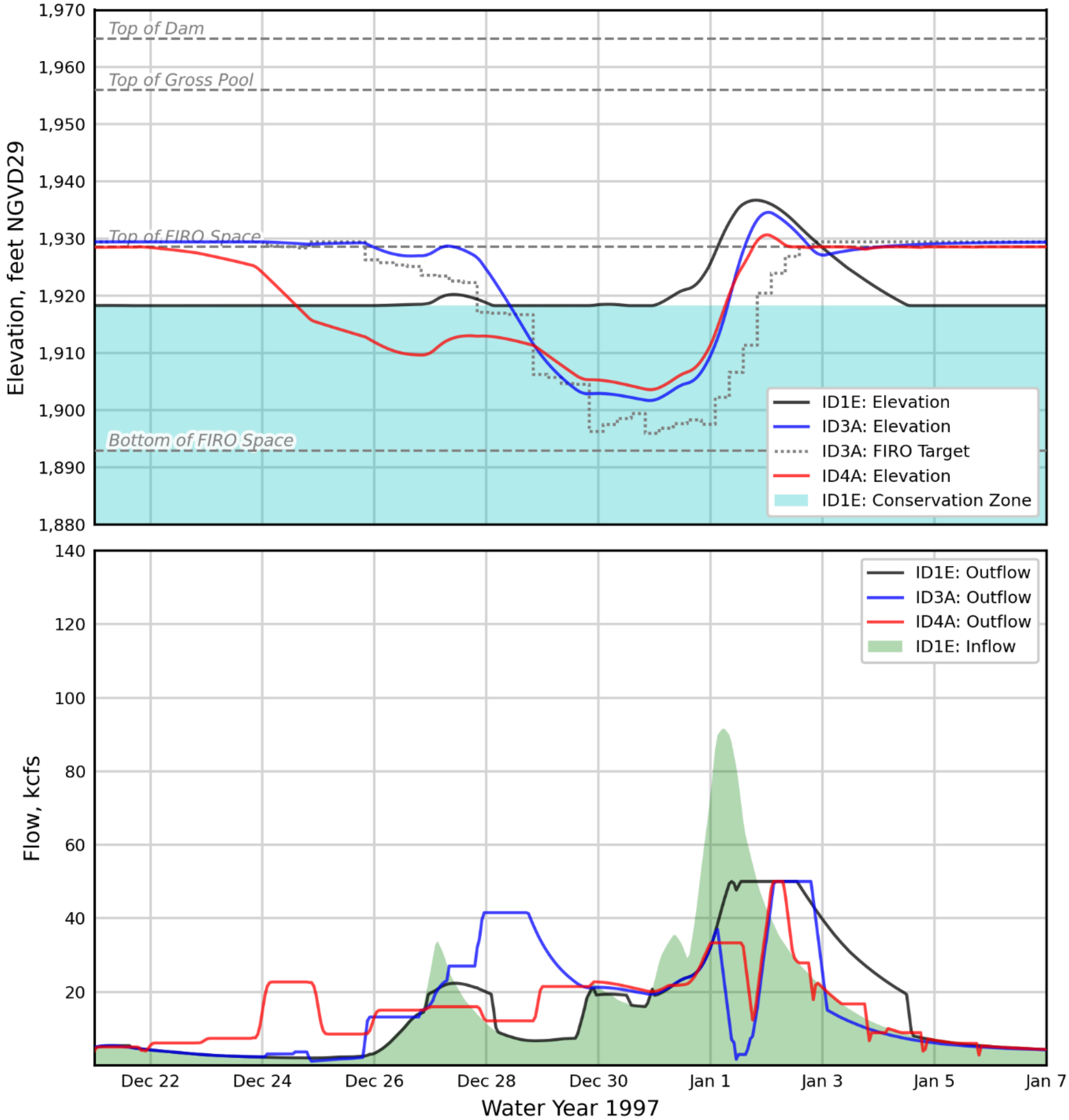
1997 90% Scaling
Downstream Control Points



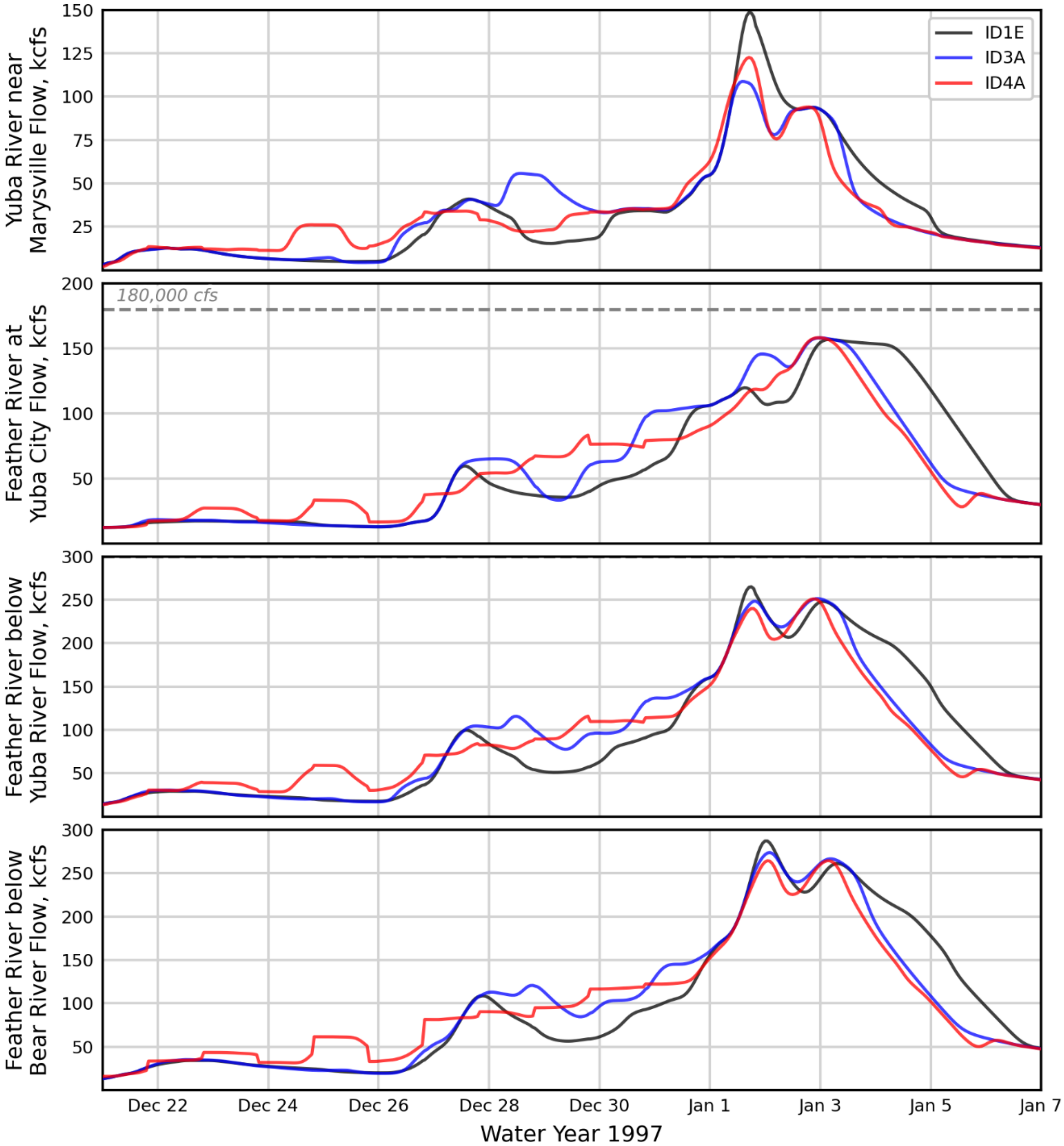
1997 90% Scaling
Oroville



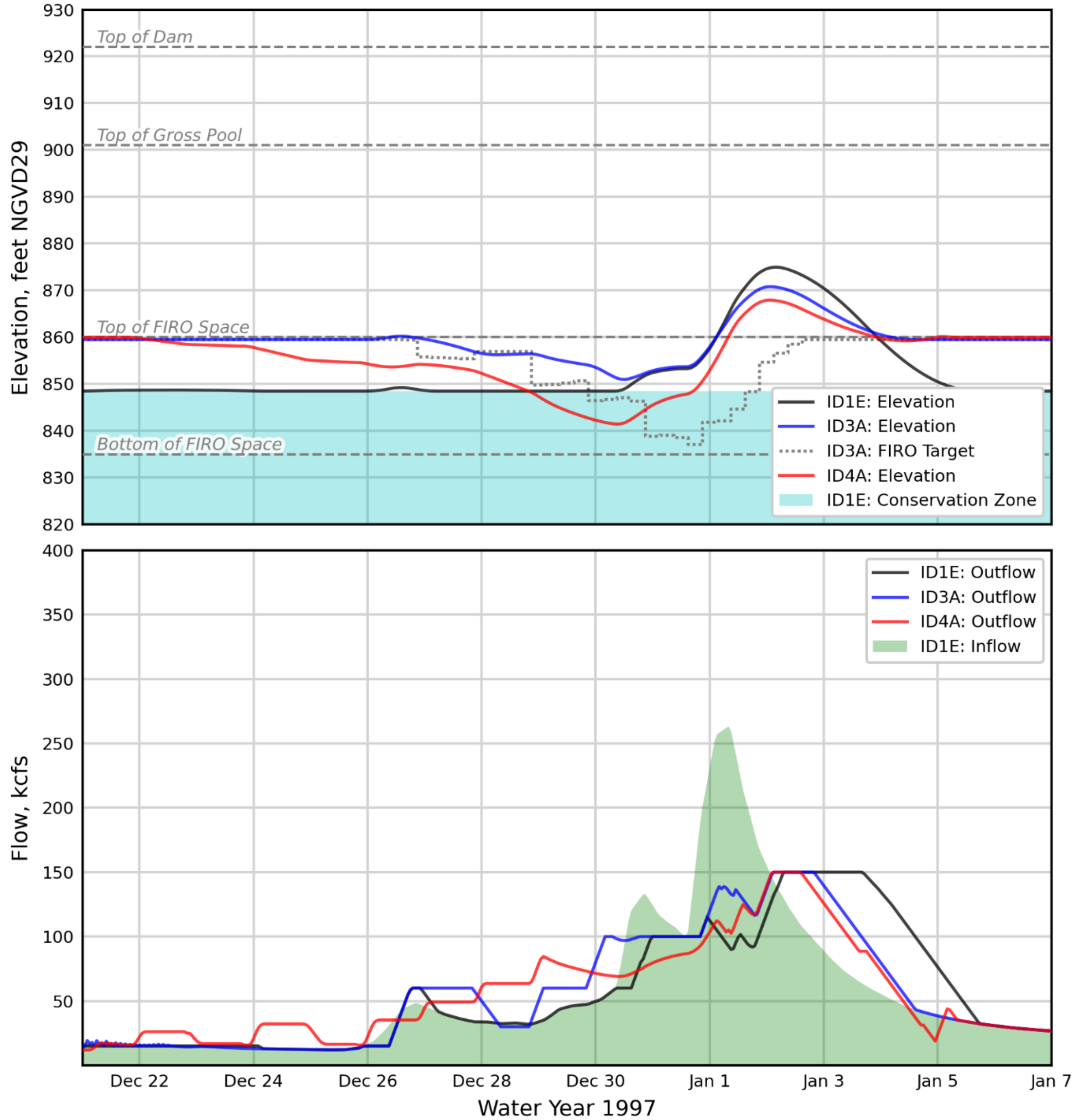
1997 90% Scaling
New Bullards Bar



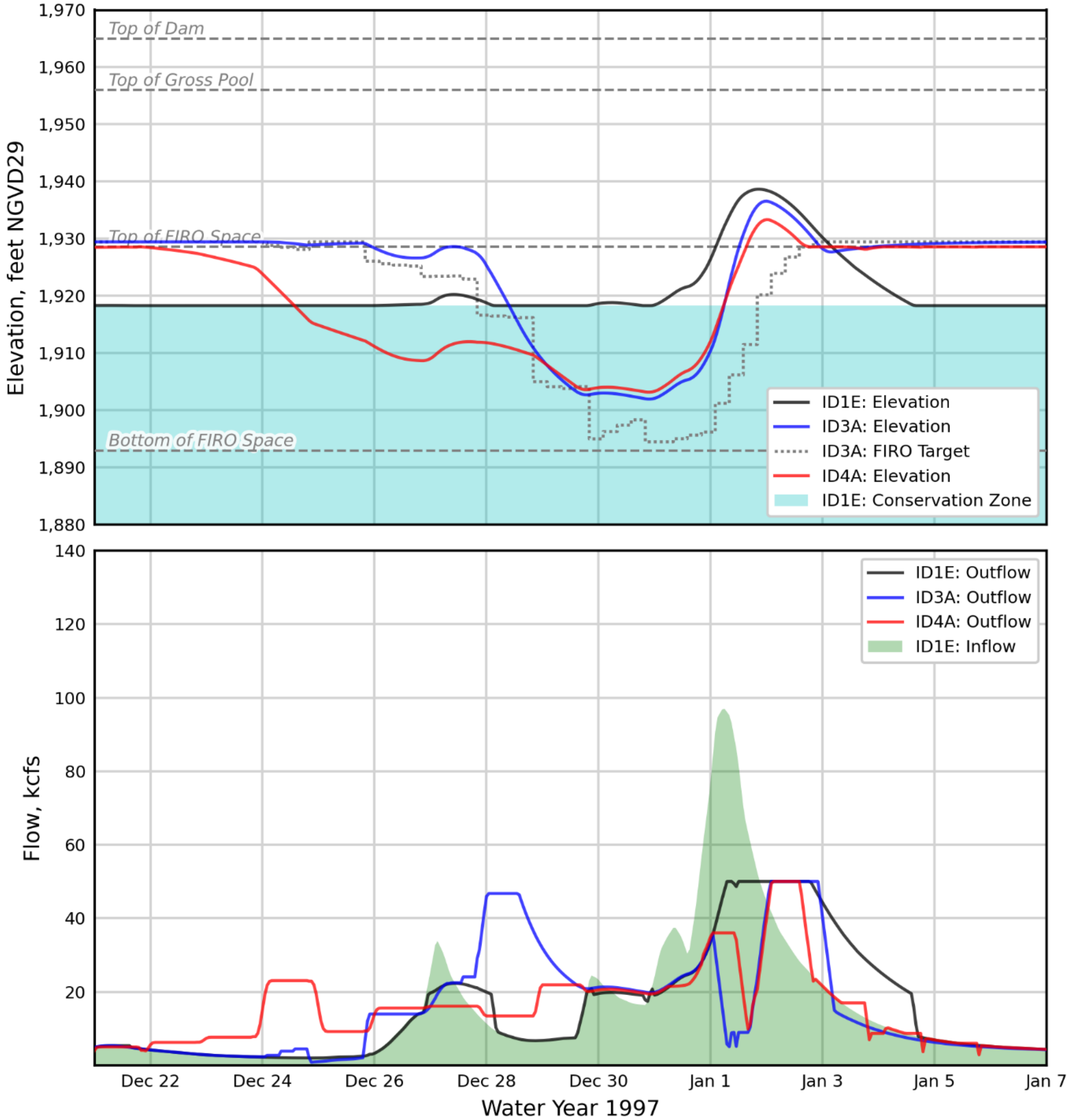
1997 92% Scaling
Downstream Control Points



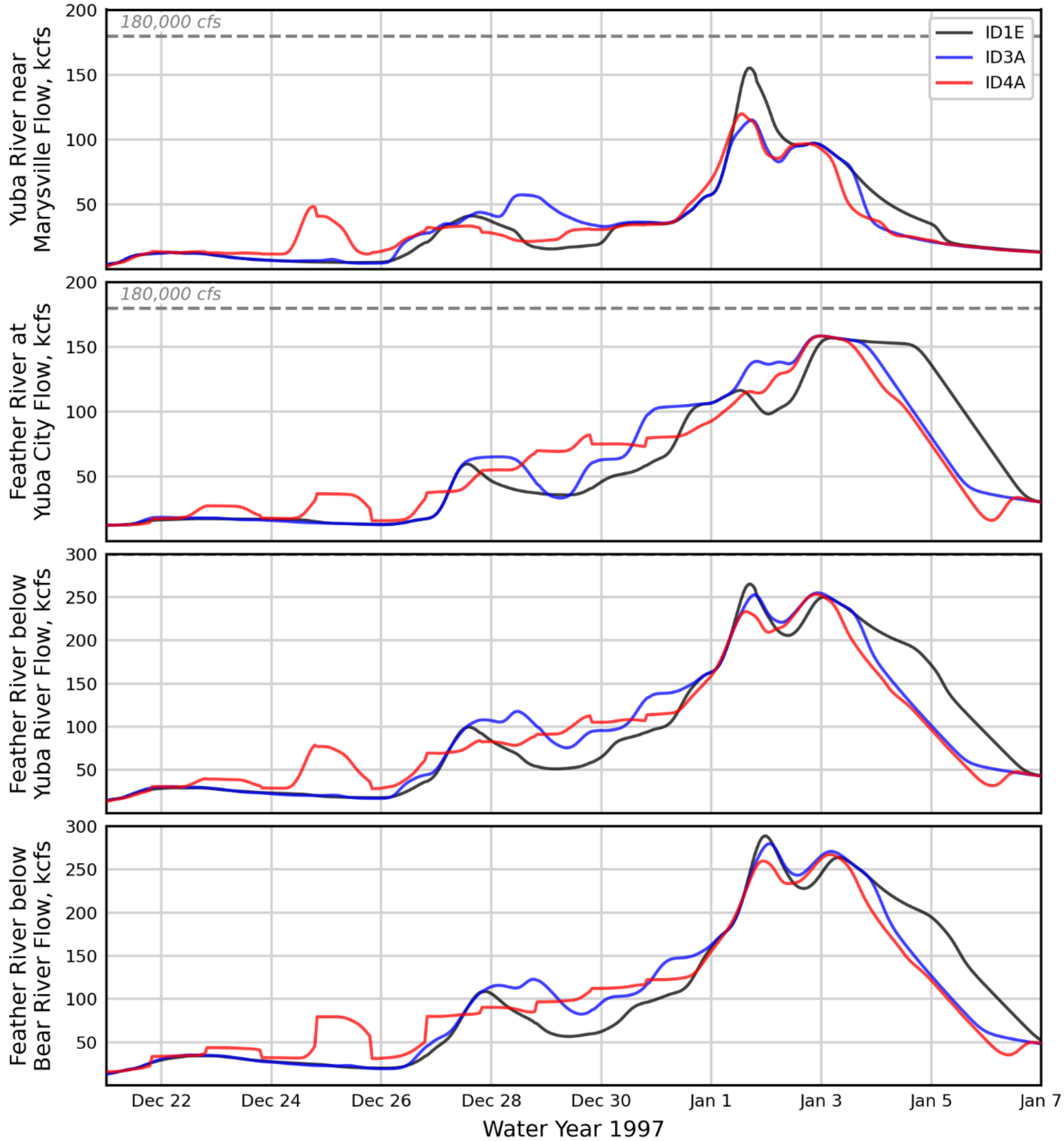
1997 92% Scaling
Oroville



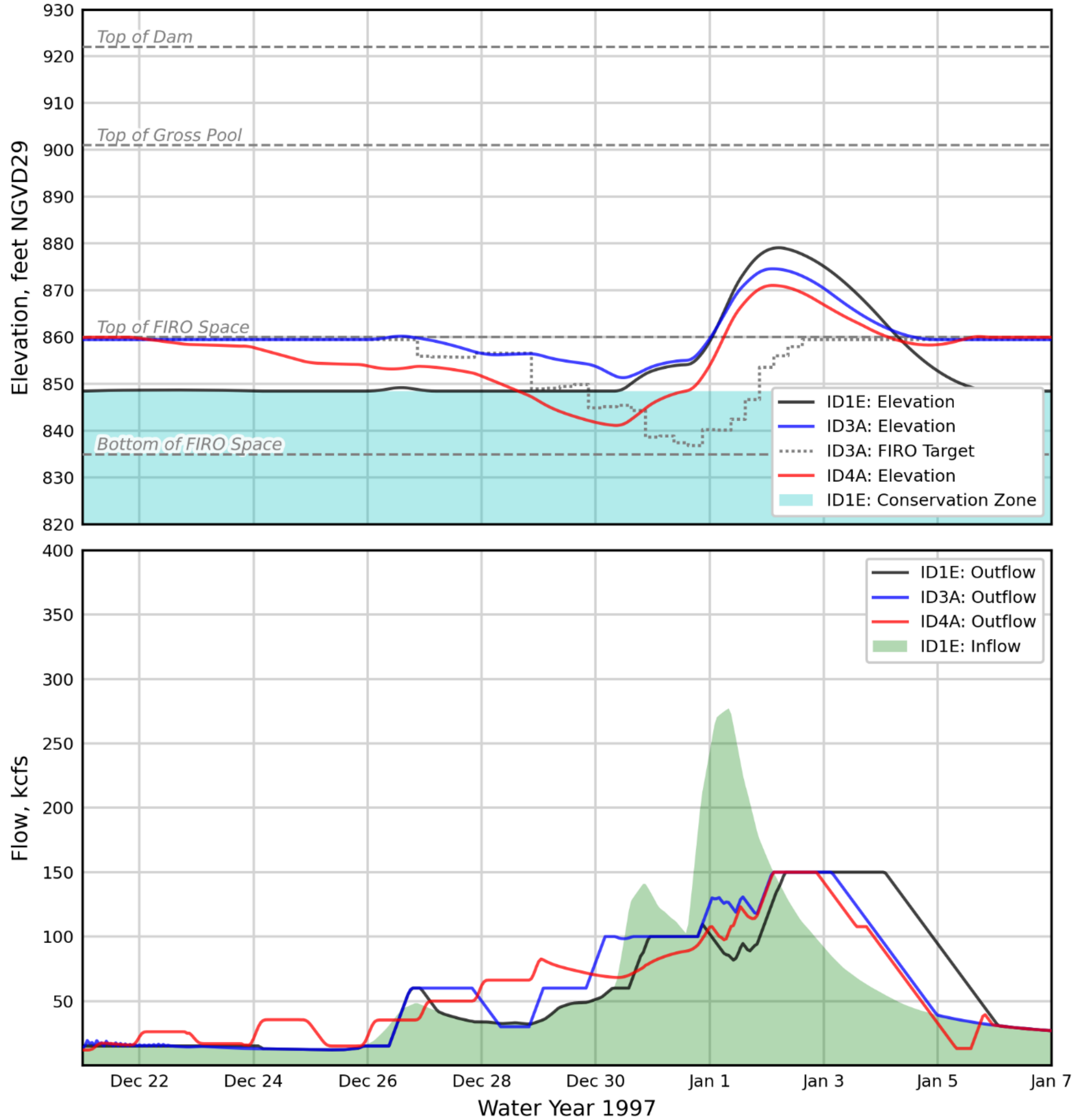
1997 92% Scaling
New Bullards Bar



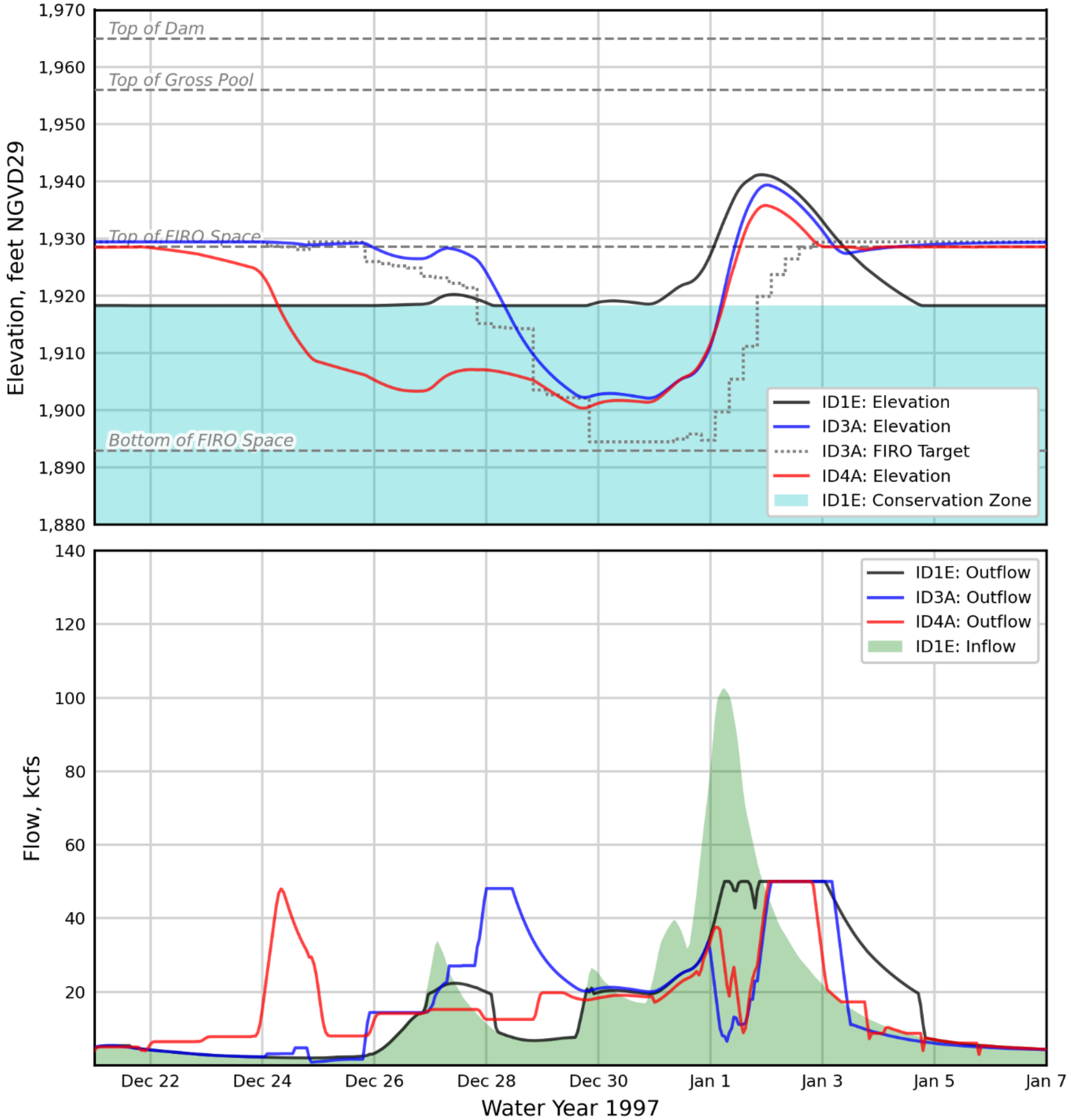
1997 94% Scaling
Downstream Control Points



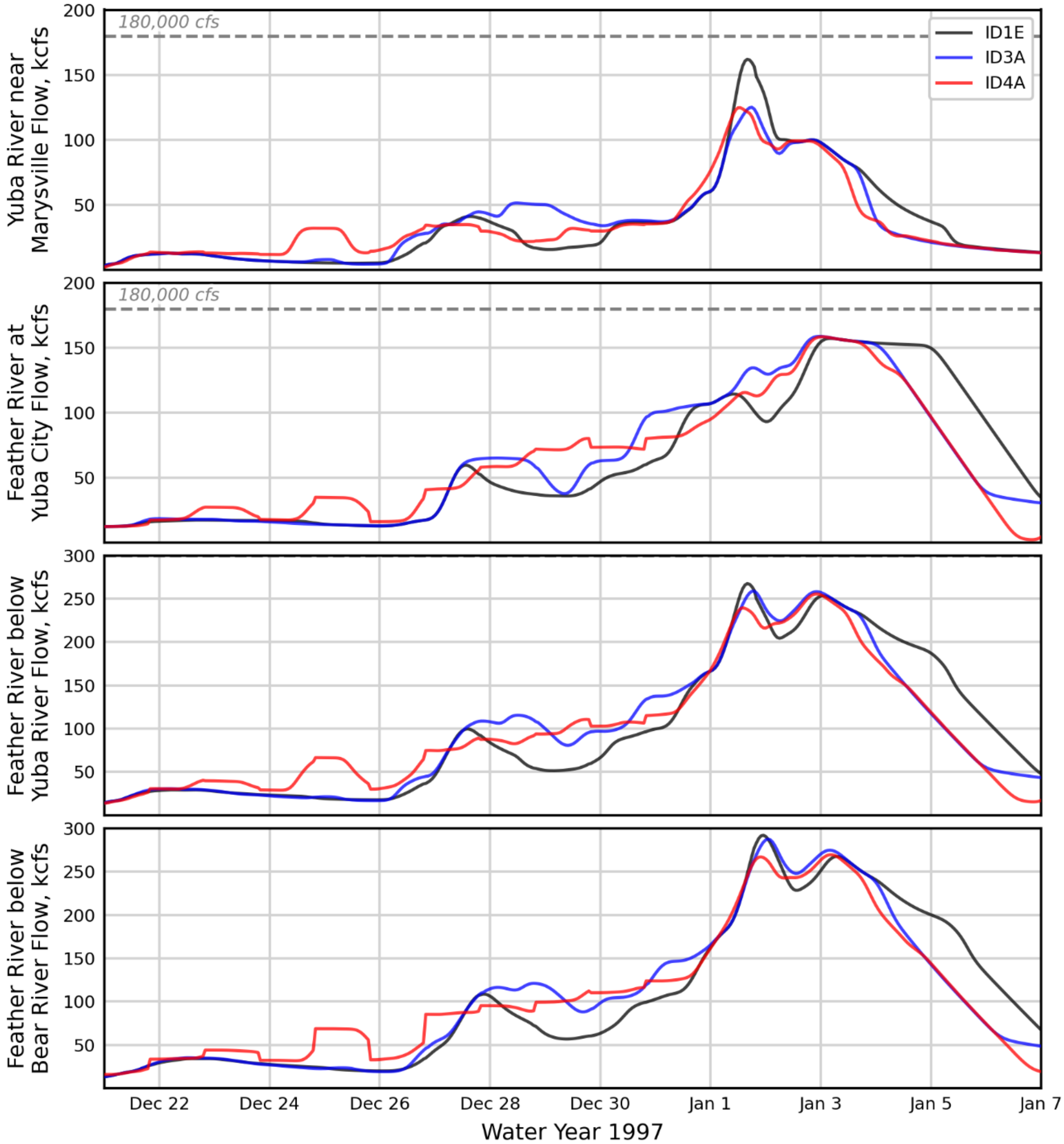
1997 94% Scaling
Oroville



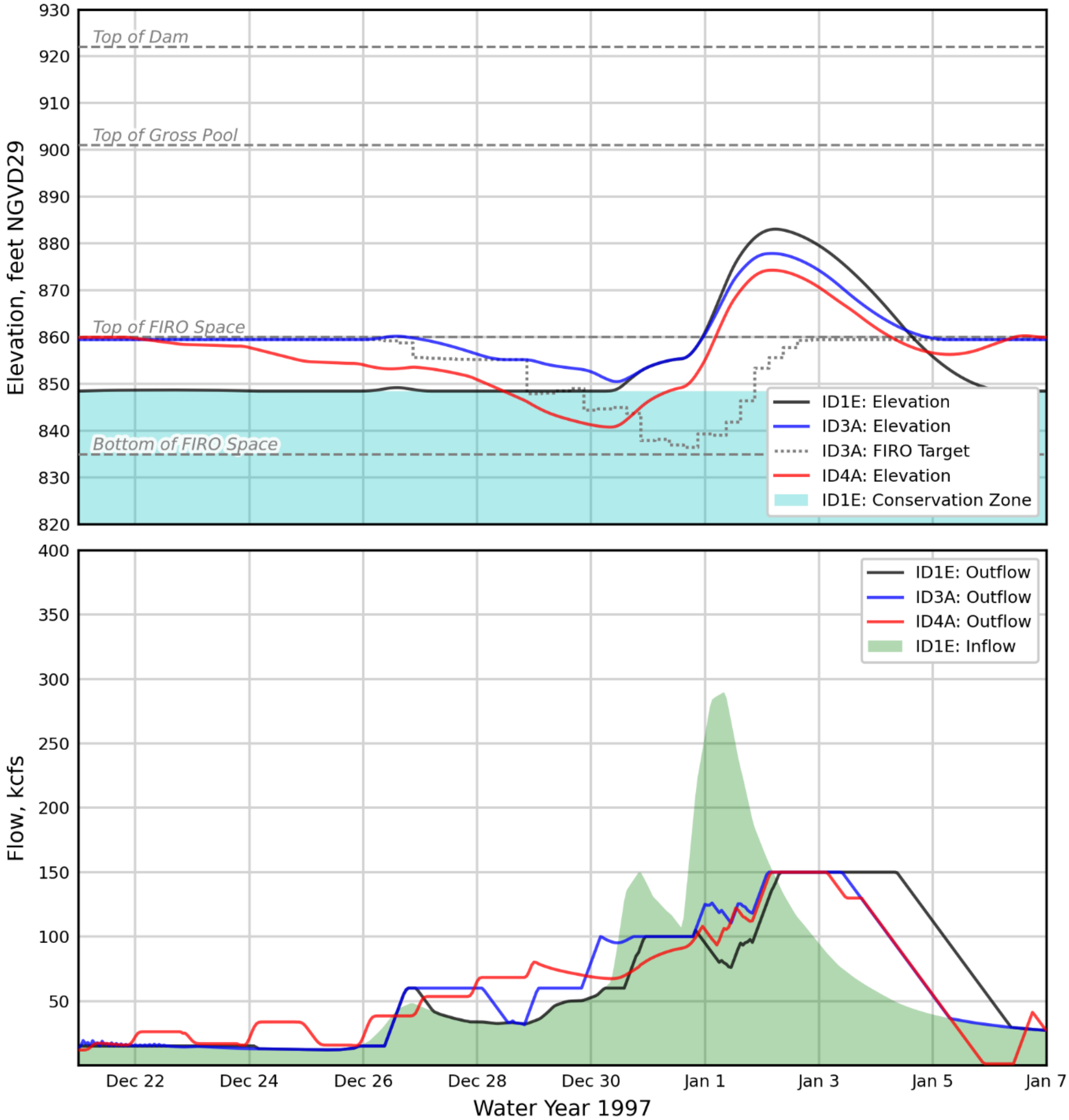
1997 94% Scaling
New Bullards Bar



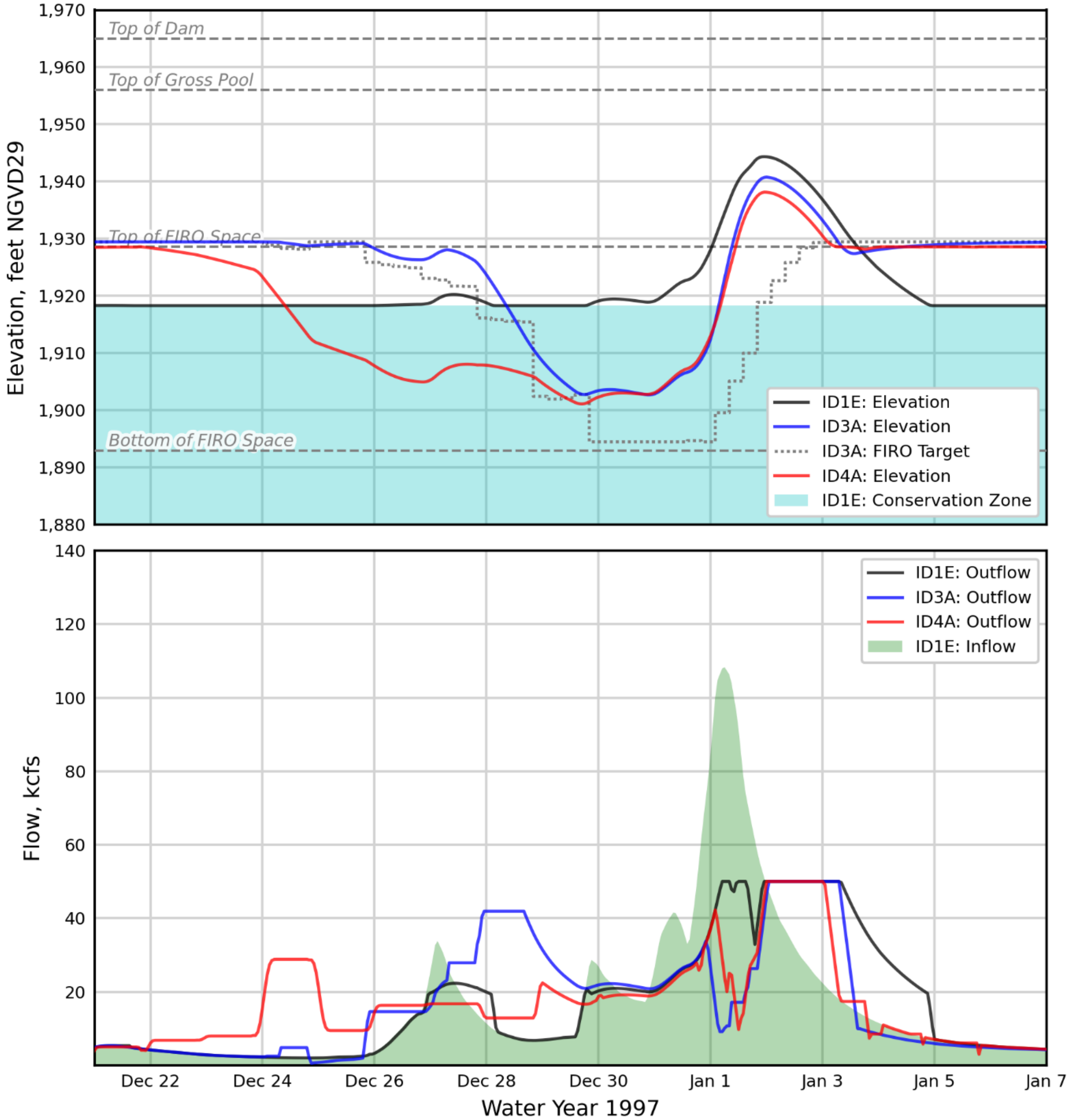
1997 96% Scaling
Downstream Control Points



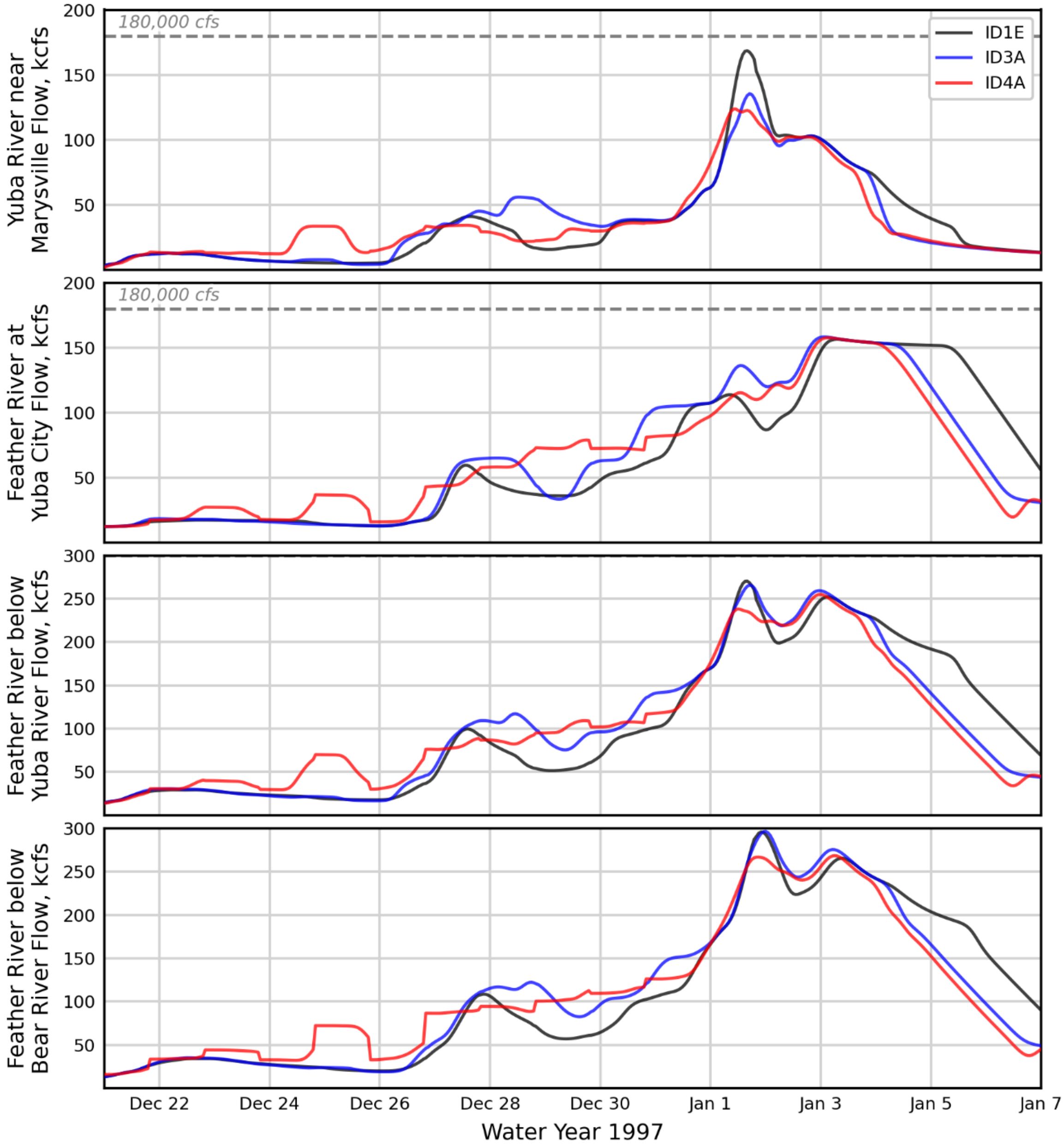
1997 96% Scaling
Oroville



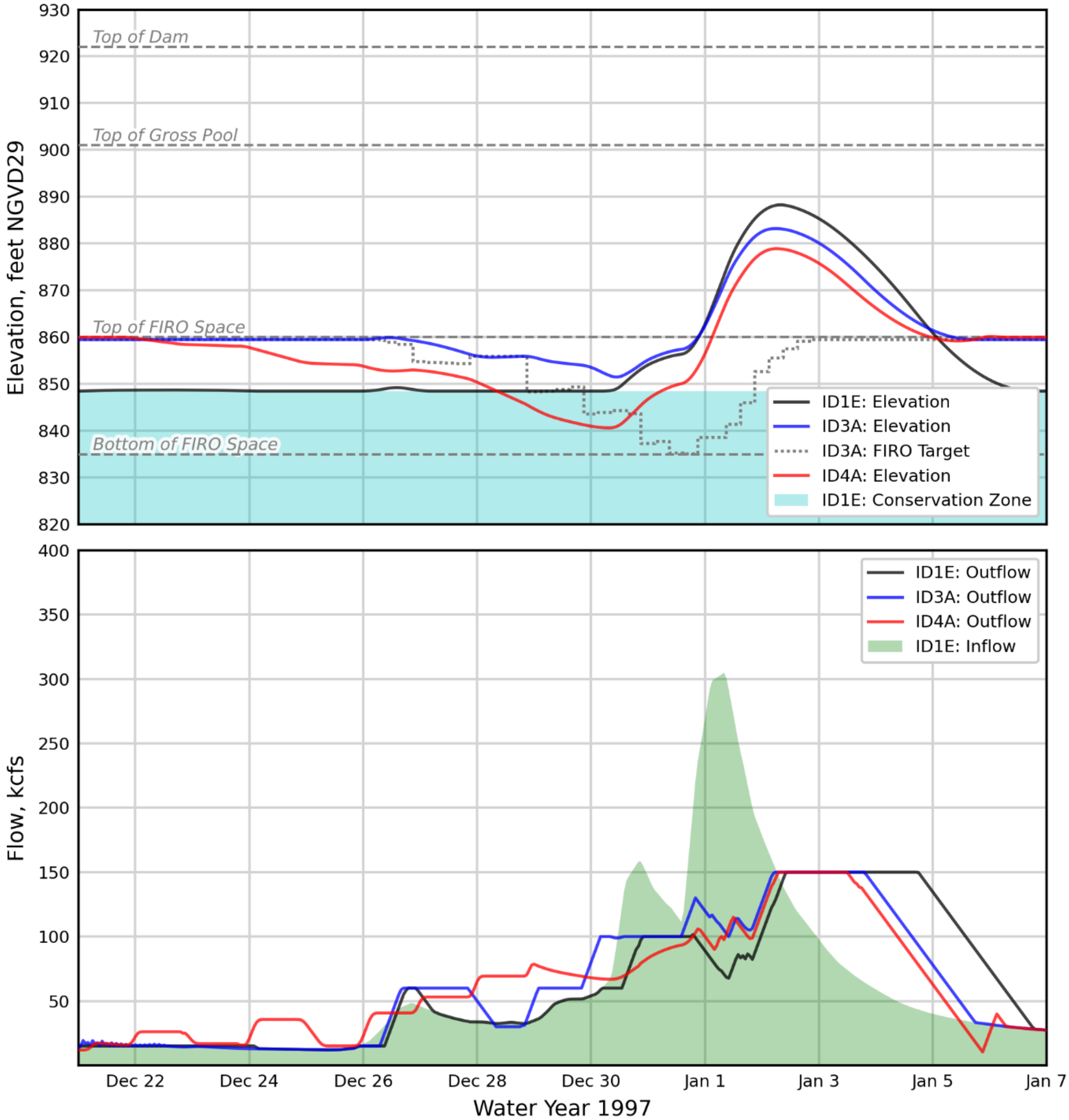
1997 96% Scaling
New Bullards Bar



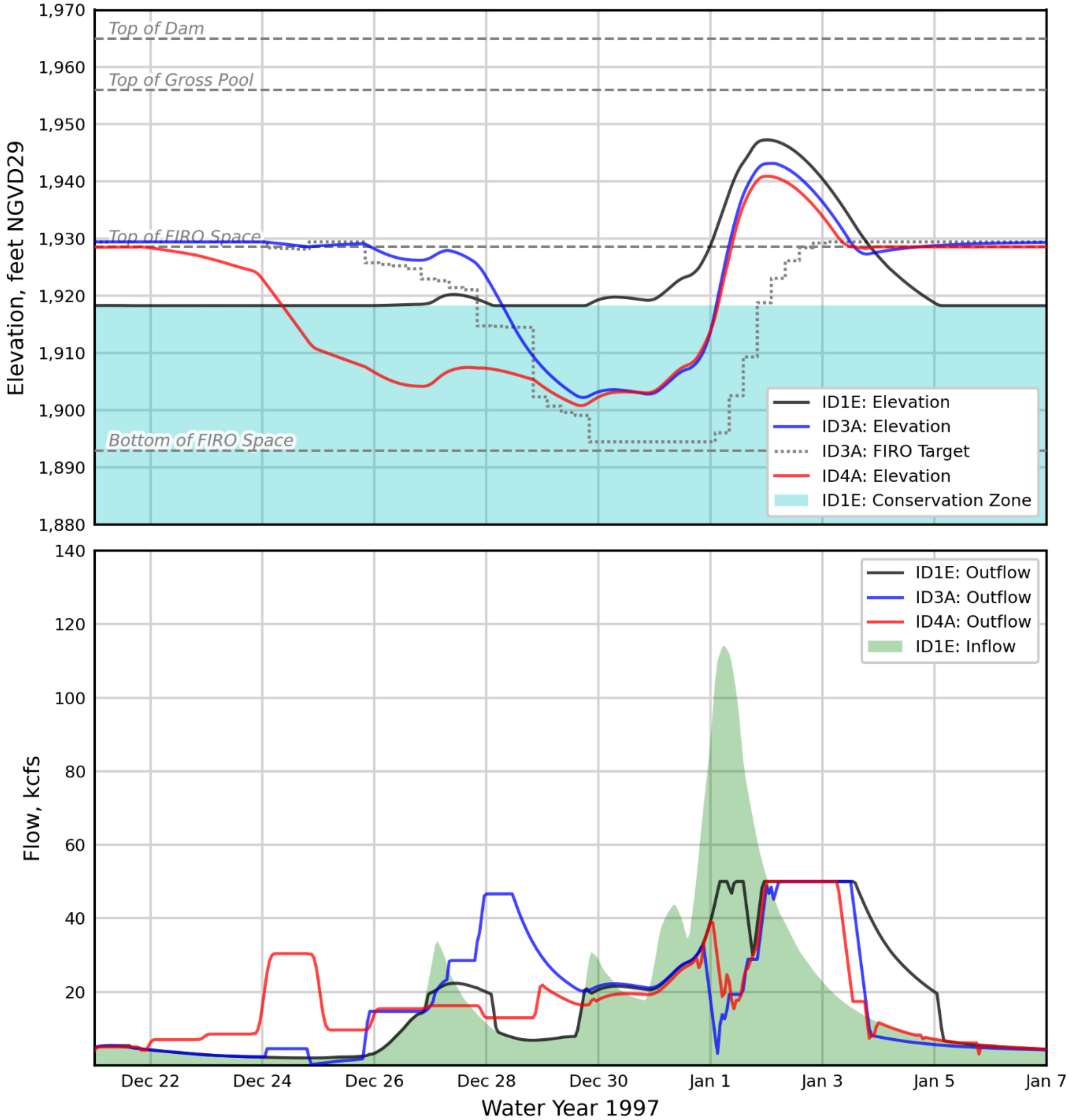
1997 98% Scaling
Downstream Control Points



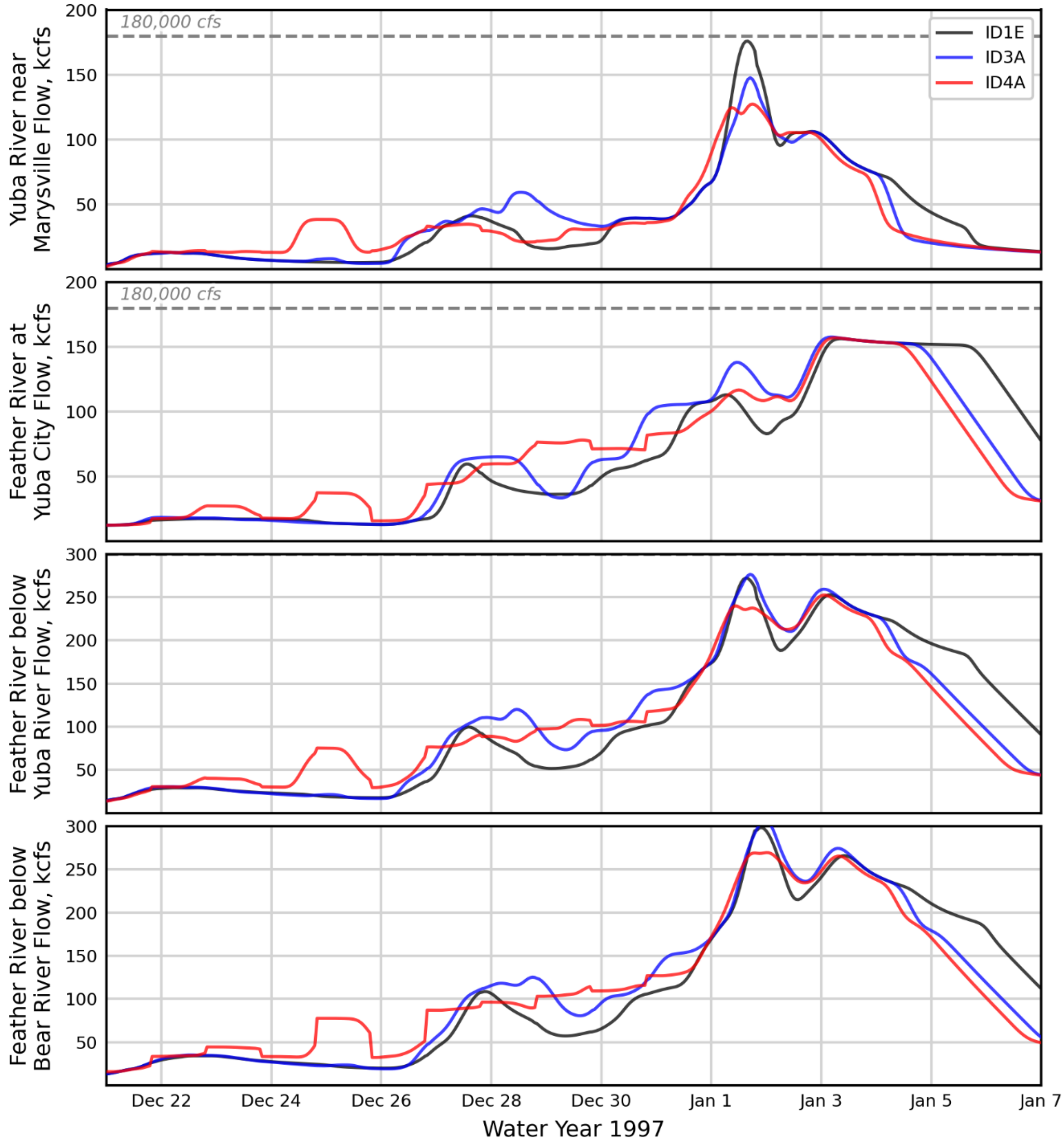
1997 98% Scaling
Oroville



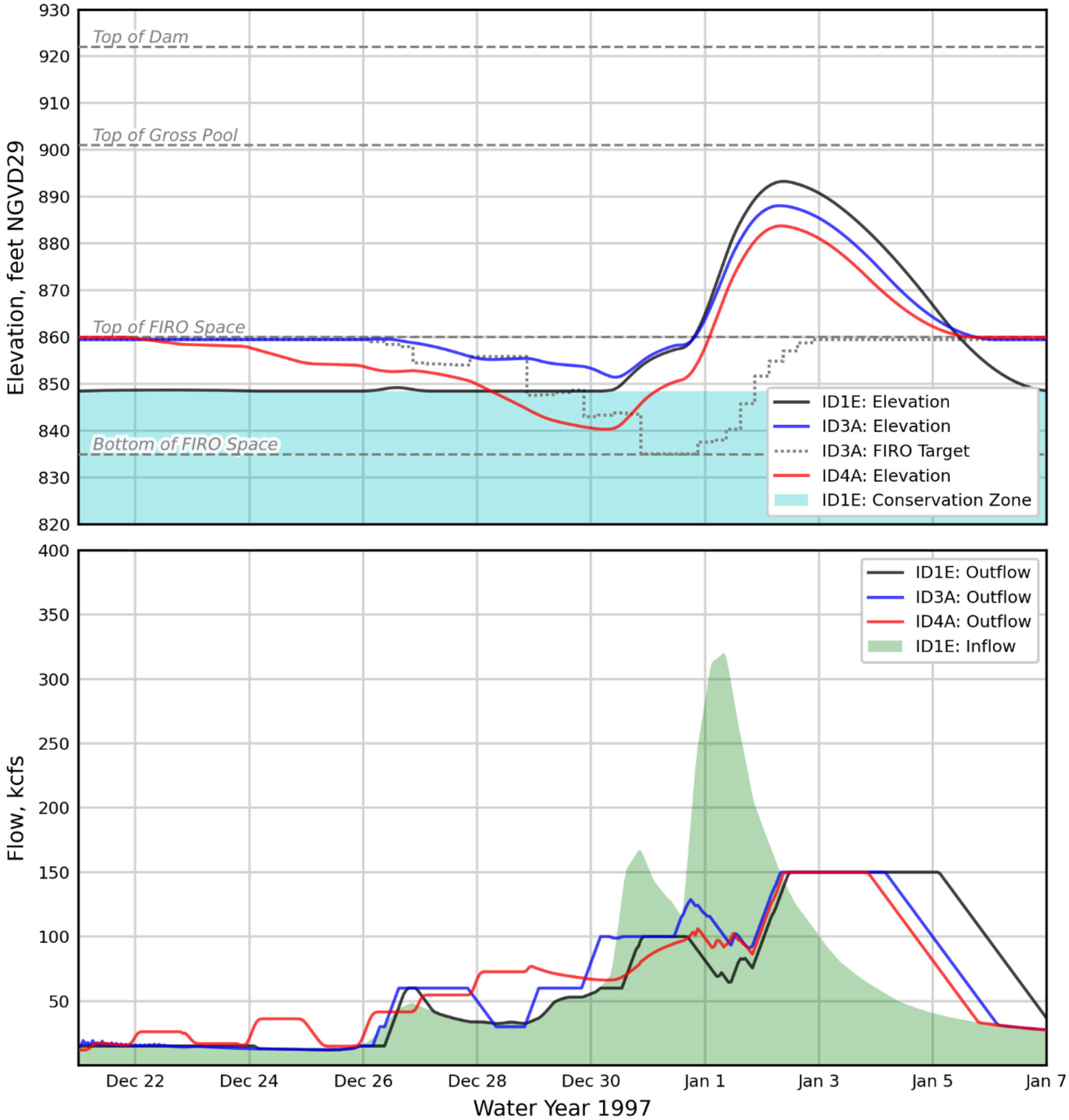
1997 98% Scaling
New Bullards Bar



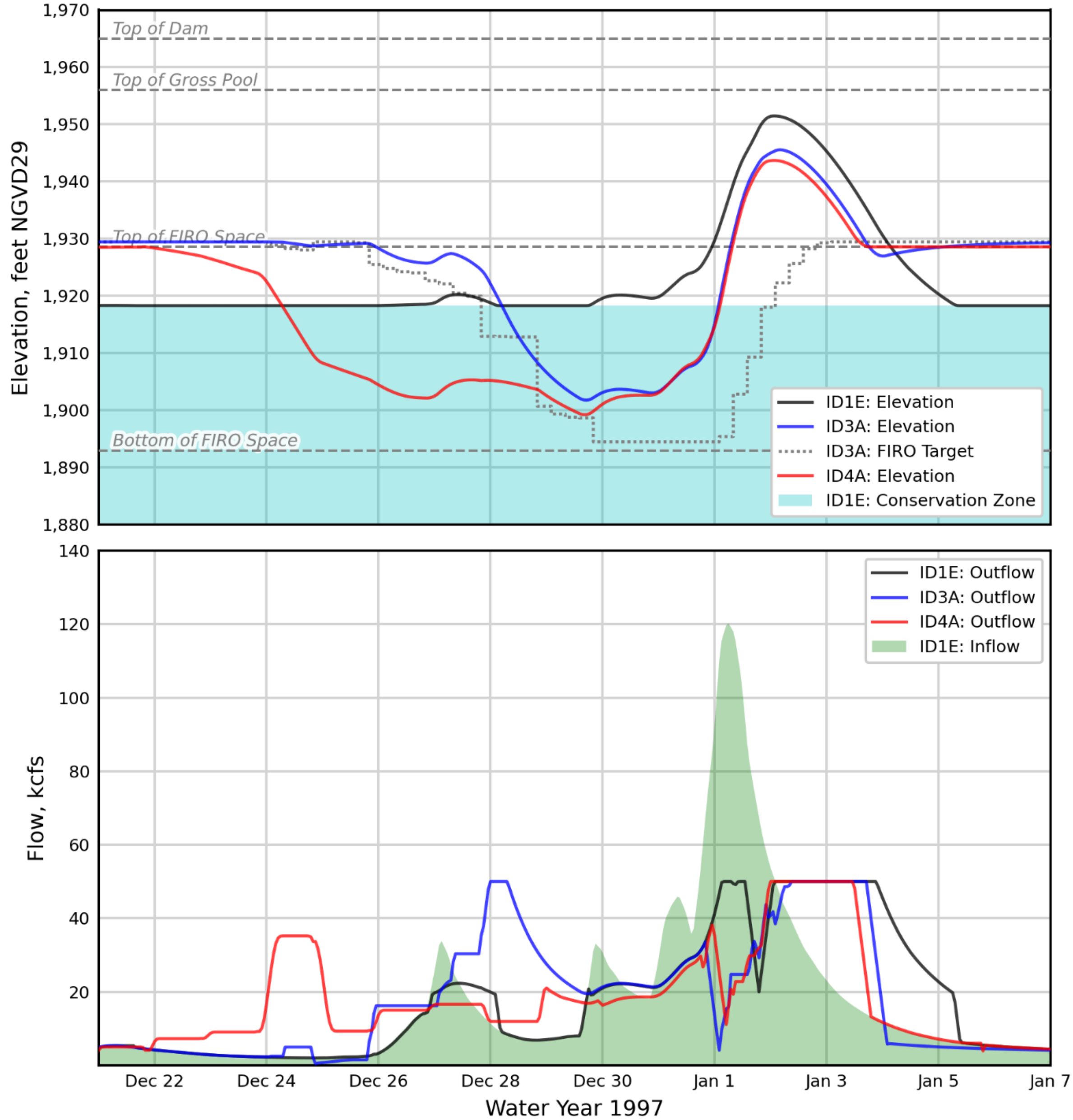
1997 100% Scaling
Downstream Control Points



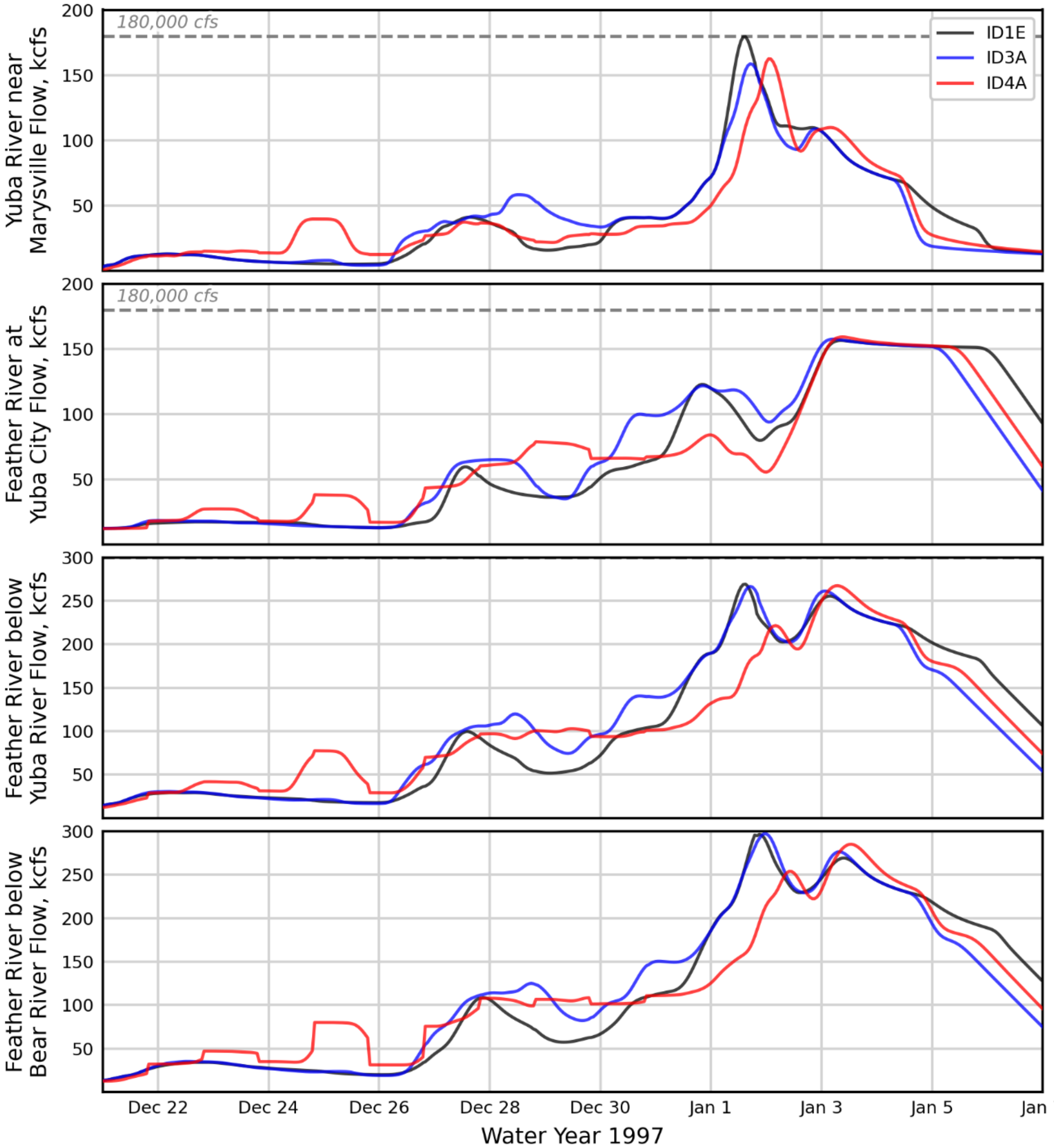
1997 100% Scaling
Oroville



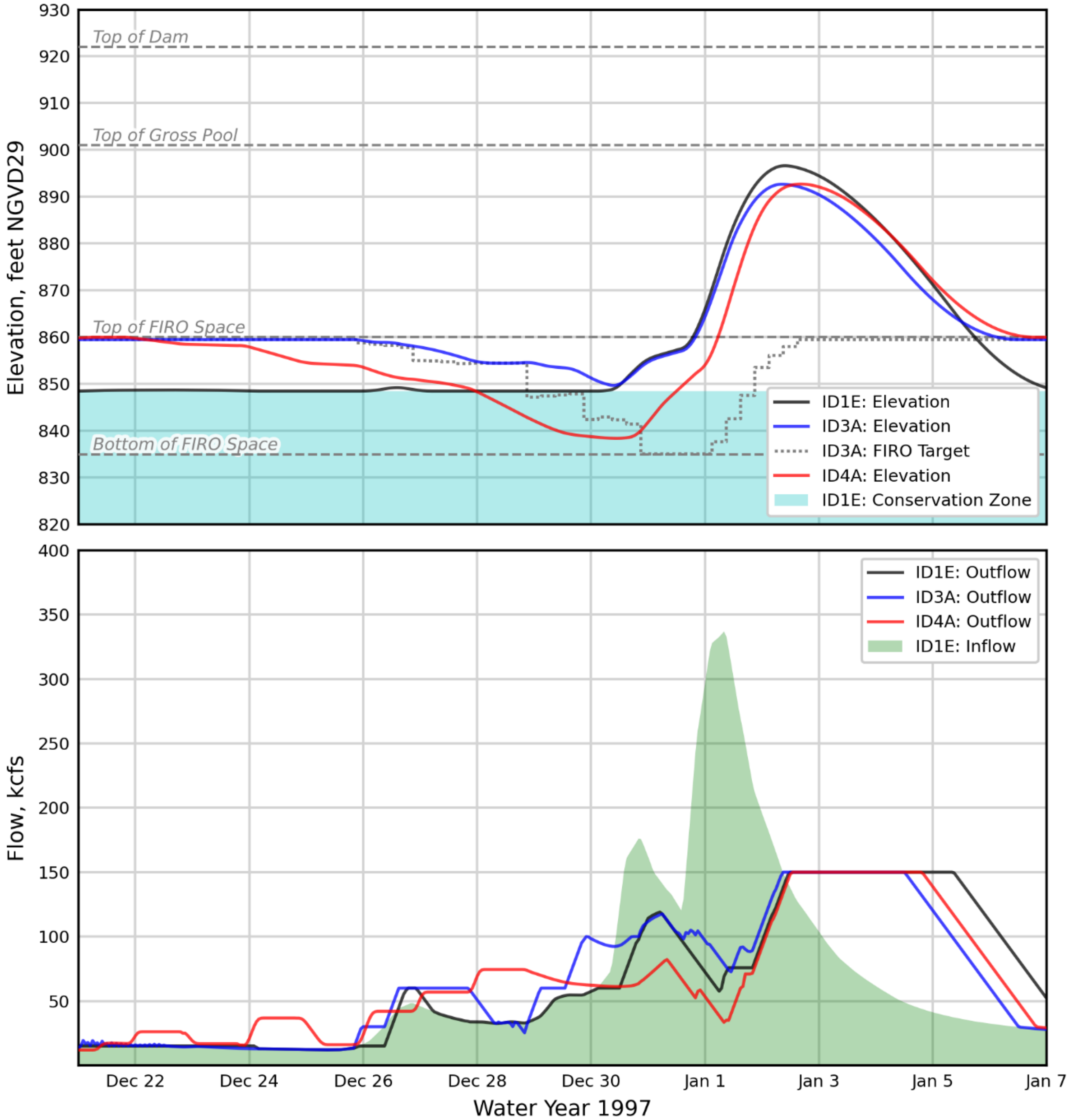
1997 100% Scaling
New Bullards Bar



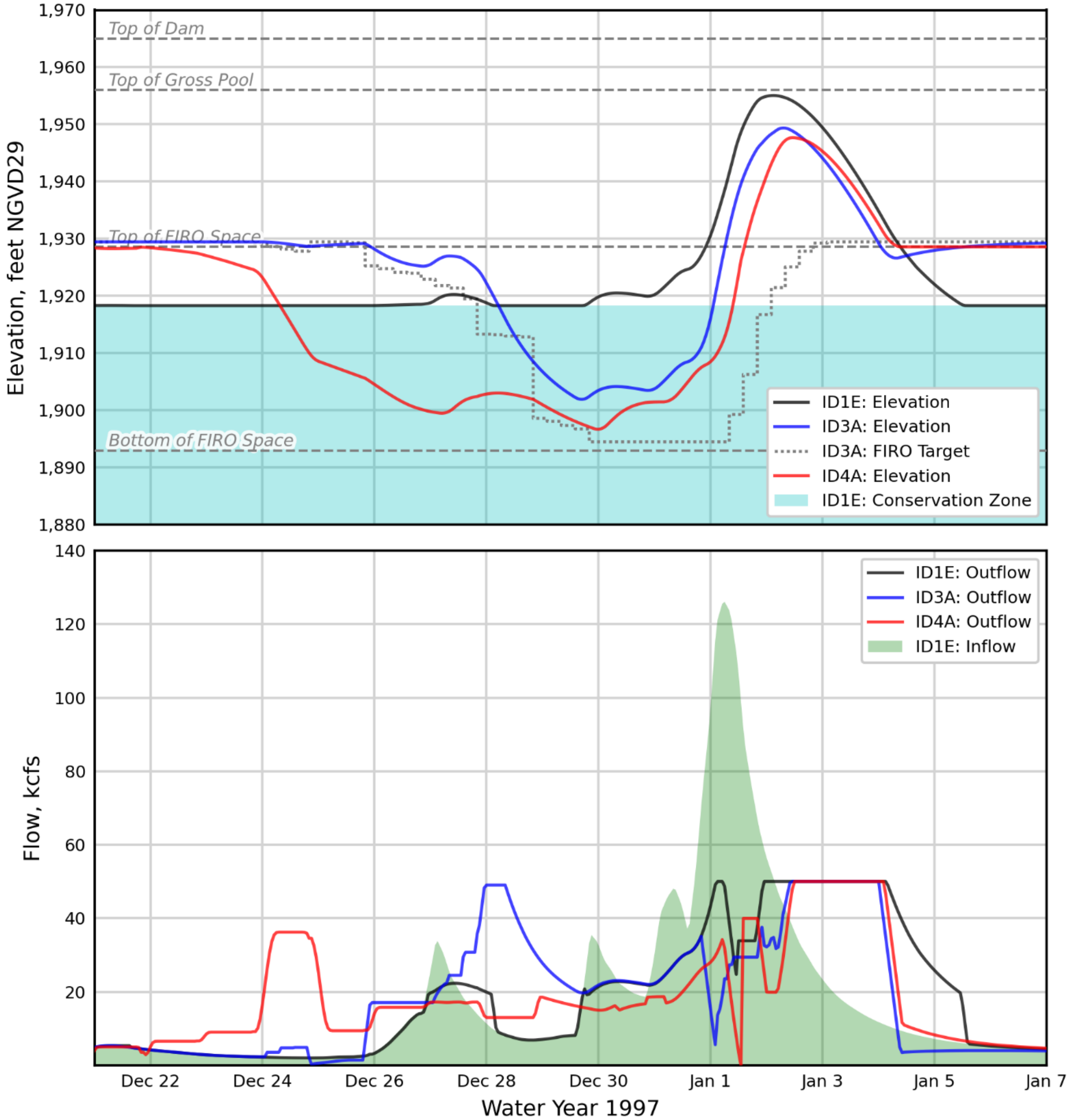
1997 102% Scaling
Downstream Control Points



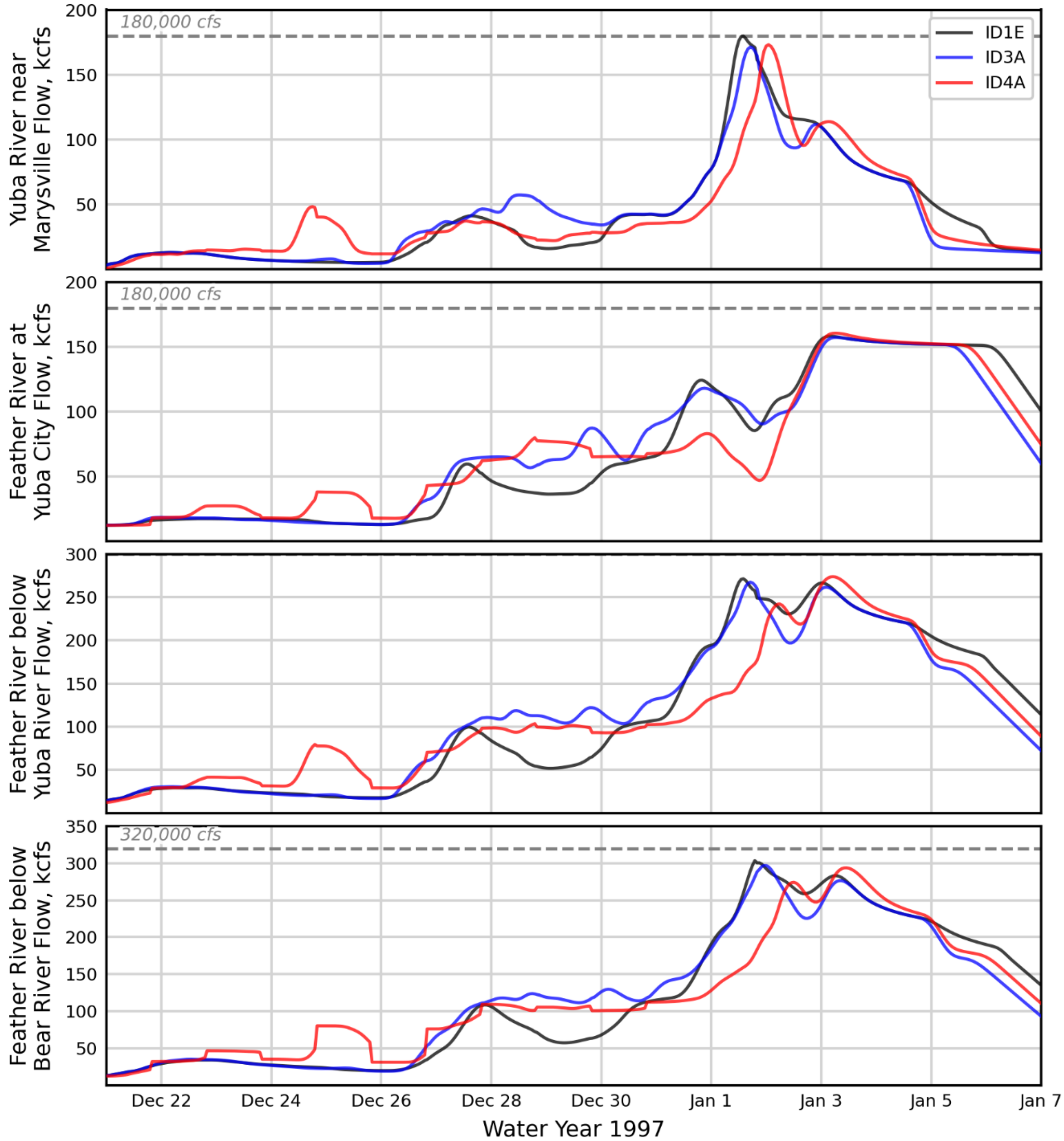
1997 102% Scaling
Oroville



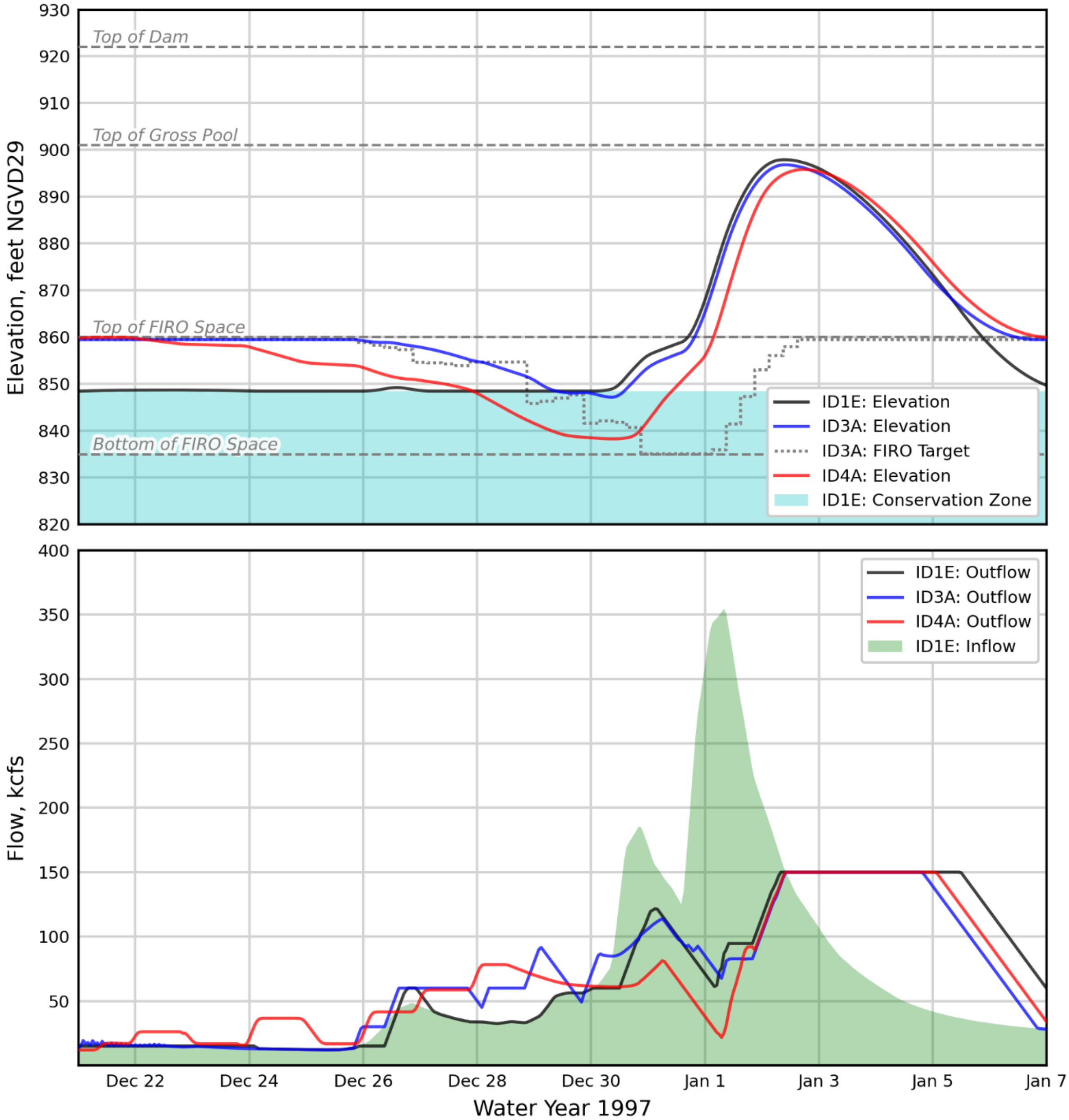
1997 102% Scaling
New Bullards Bar



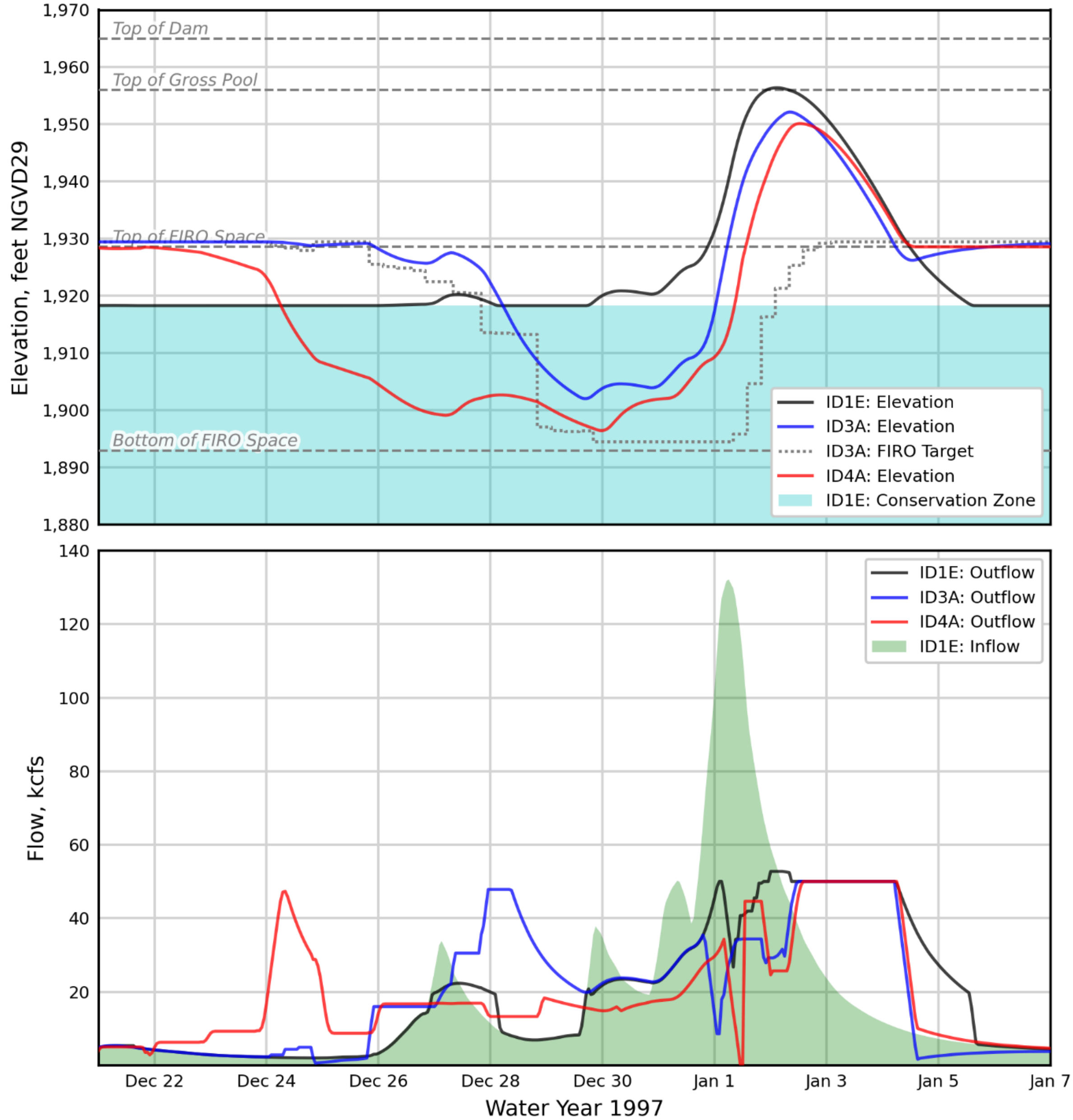
1997 104% Scaling
Downstream Control Points



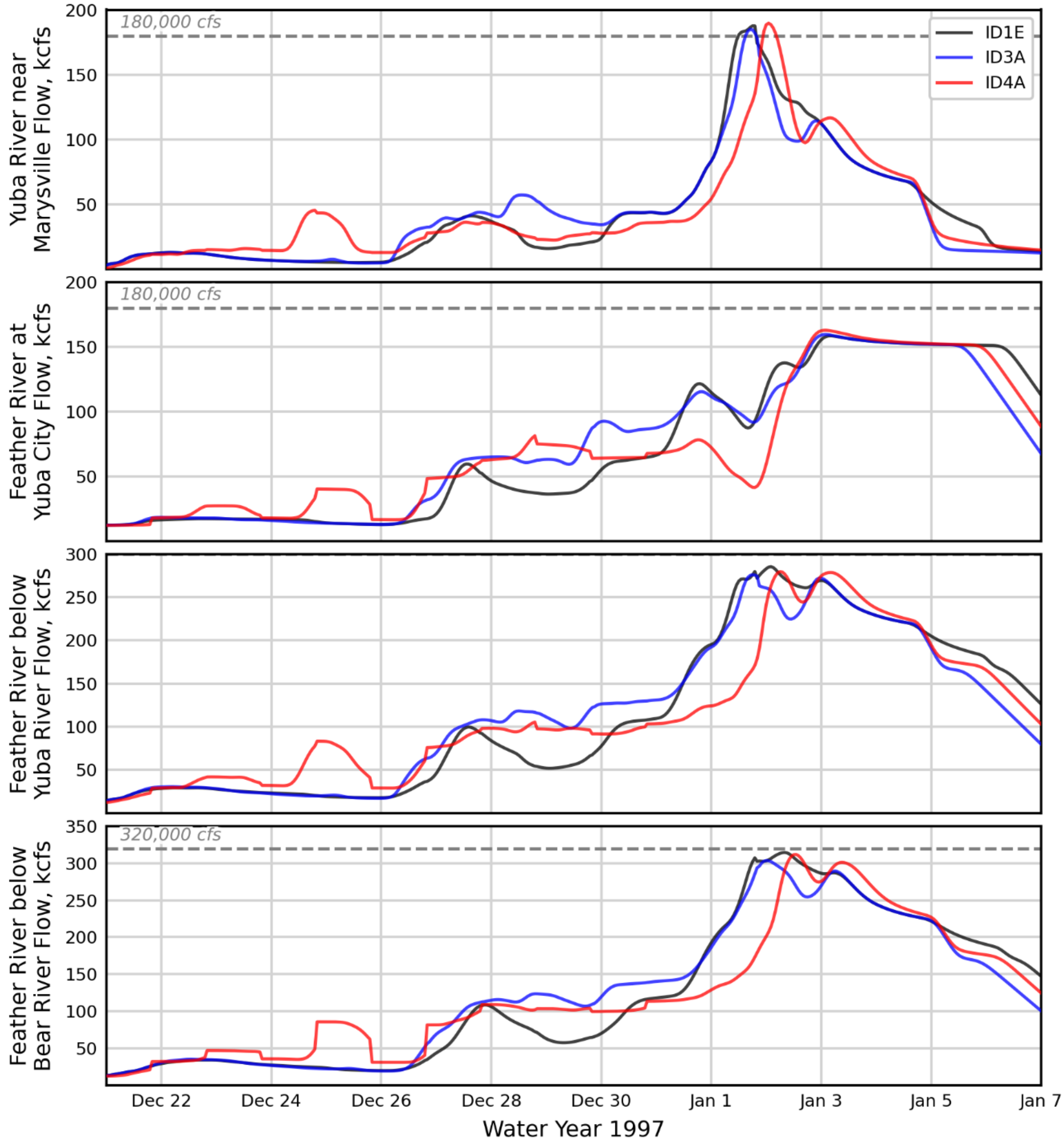
1997 104% Scaling
Oroville



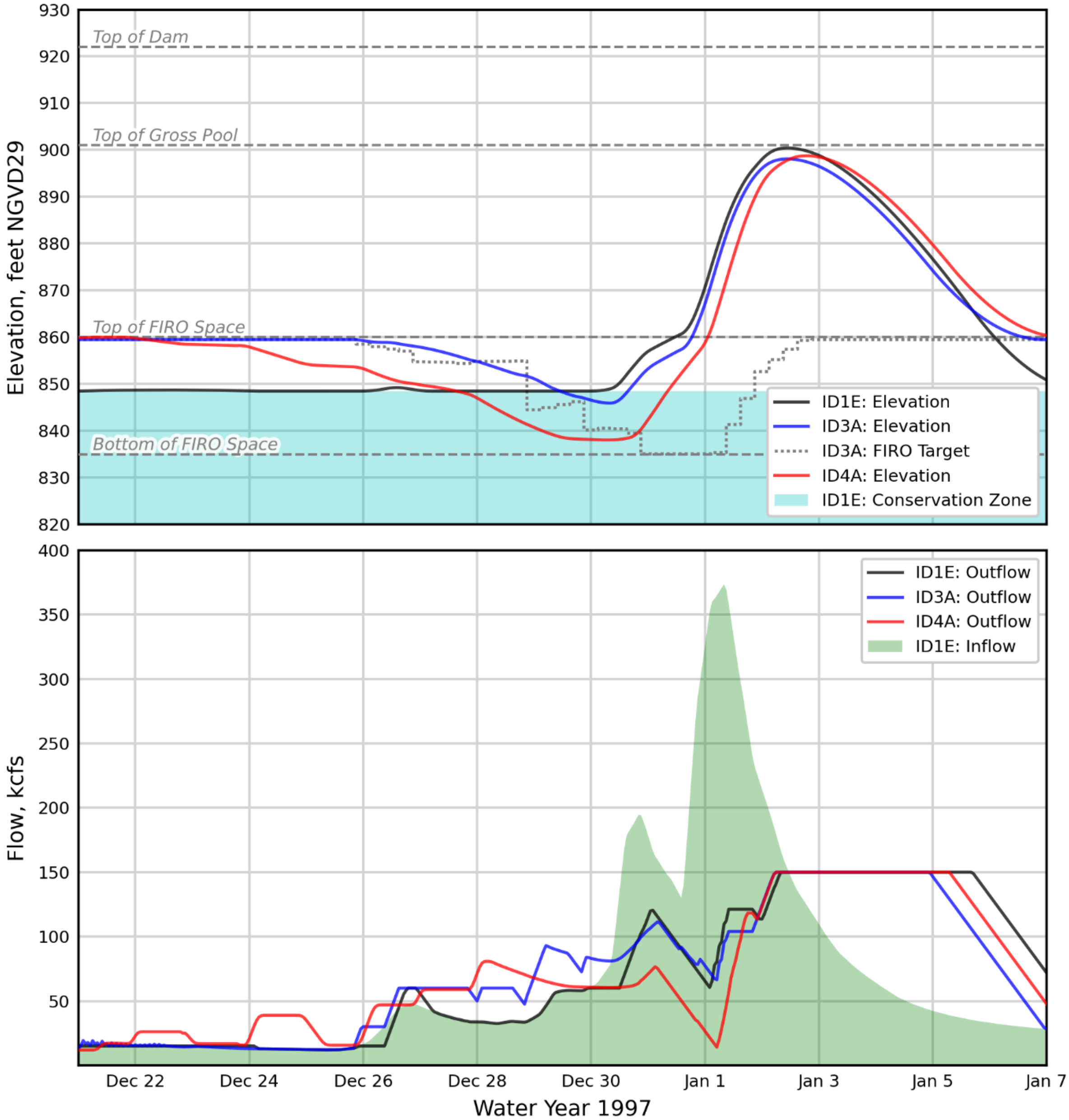
1997 104% Scaling
New Bullards Bar



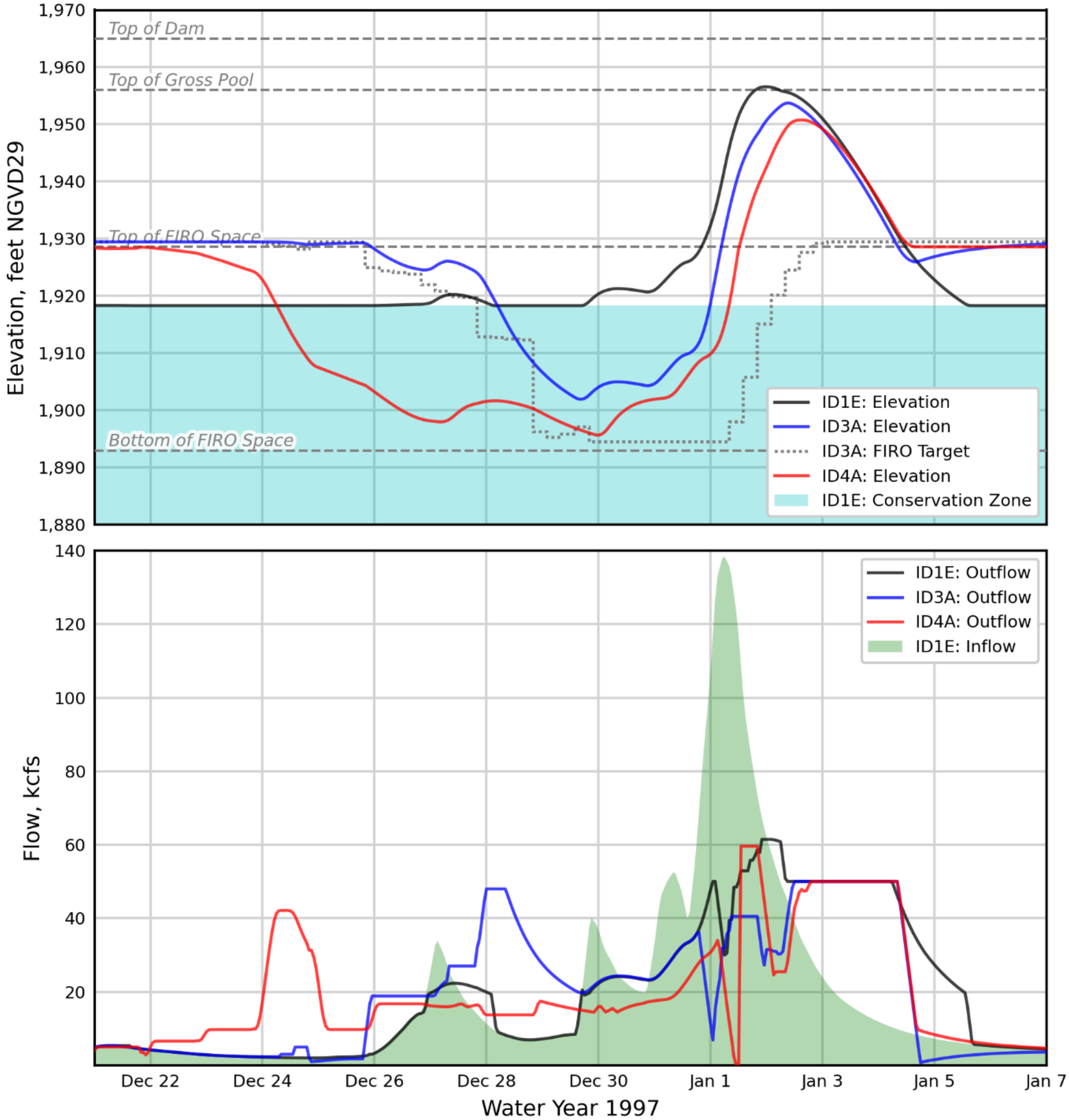
1997 106% Scaling
Downstream Control Points



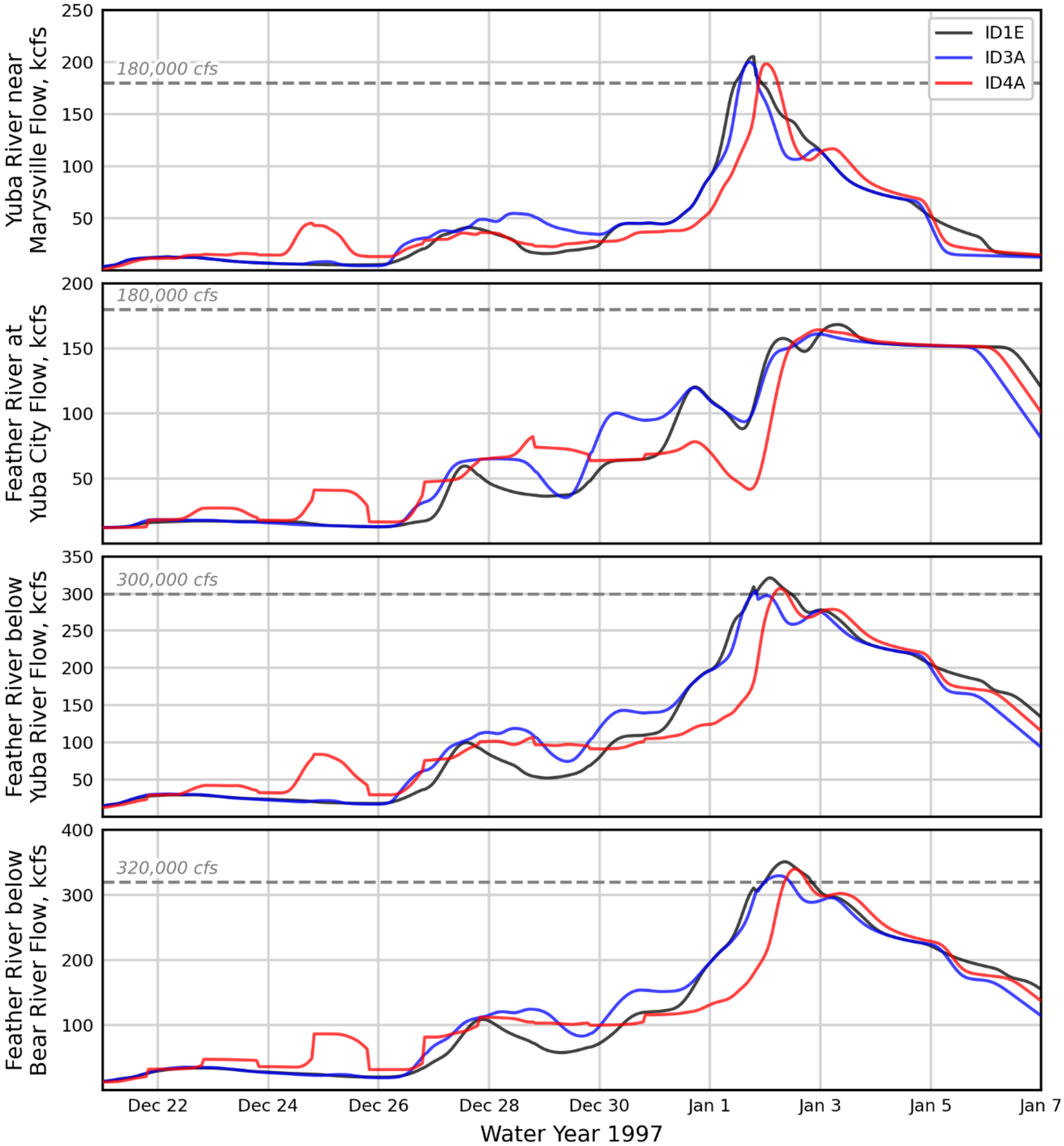
1997 106% Scaling
Oroville



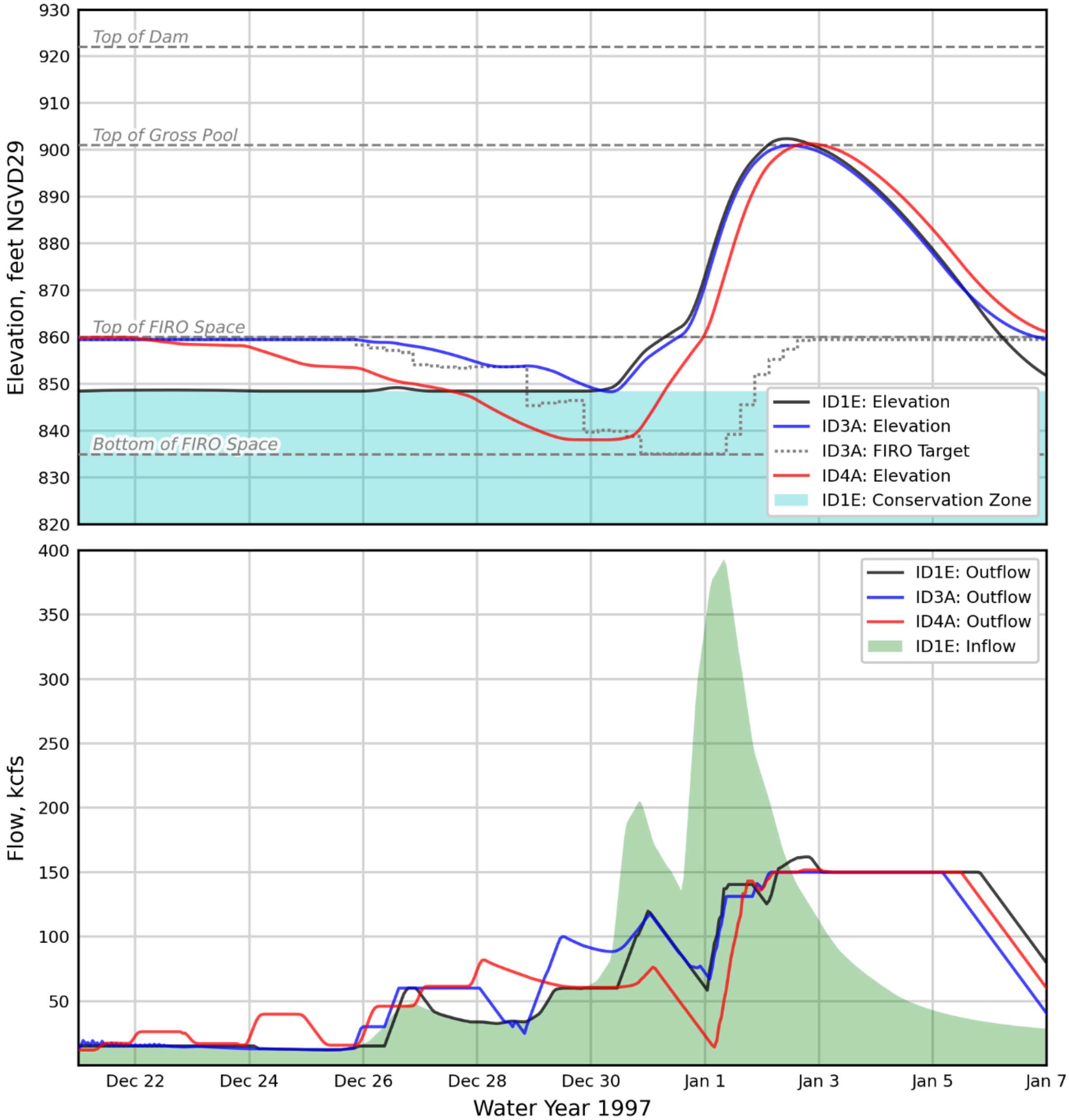
1997 106% Scaling
New Bullards Bar



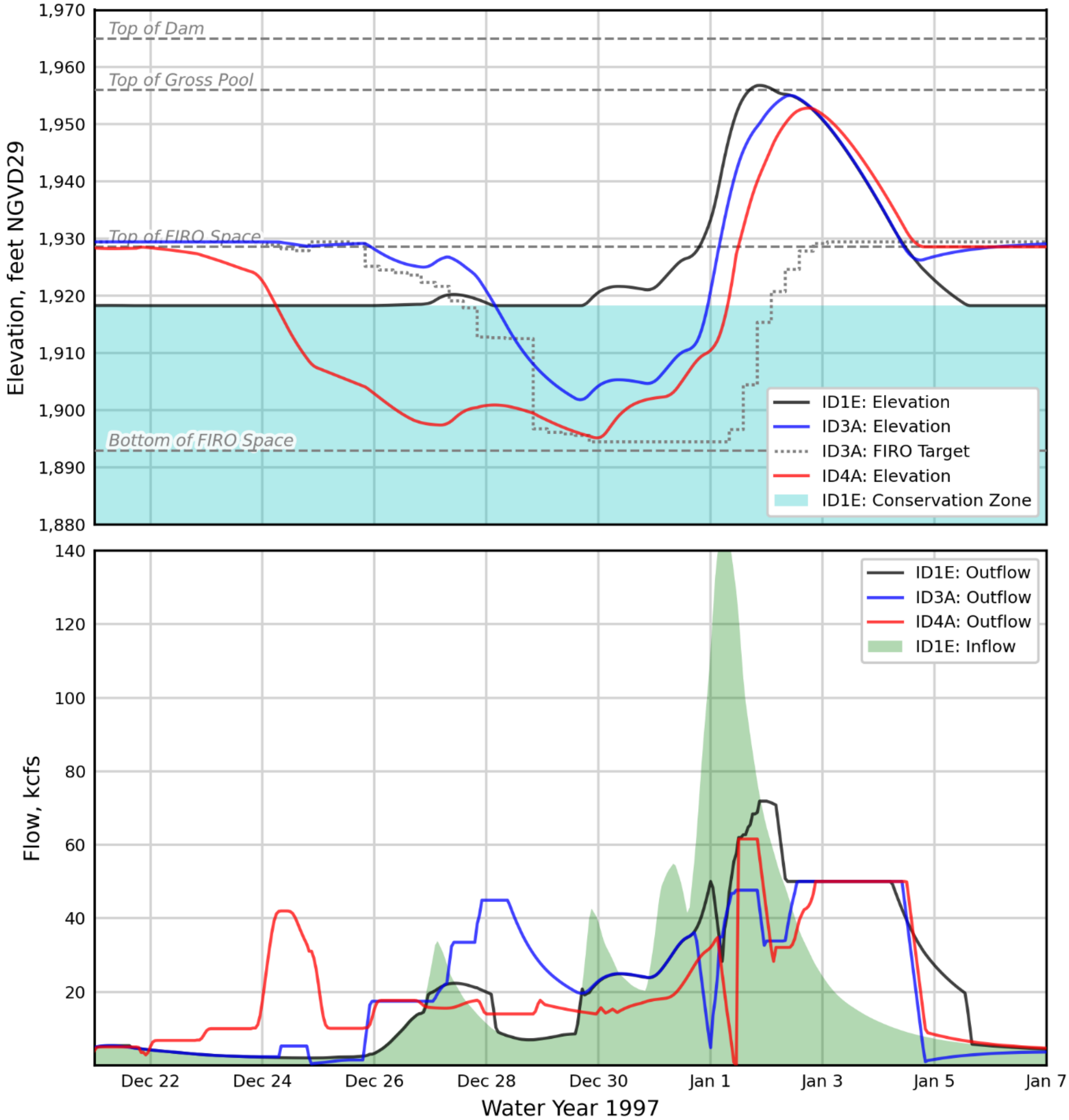
1997 108% Scaling
Downstream Control Points



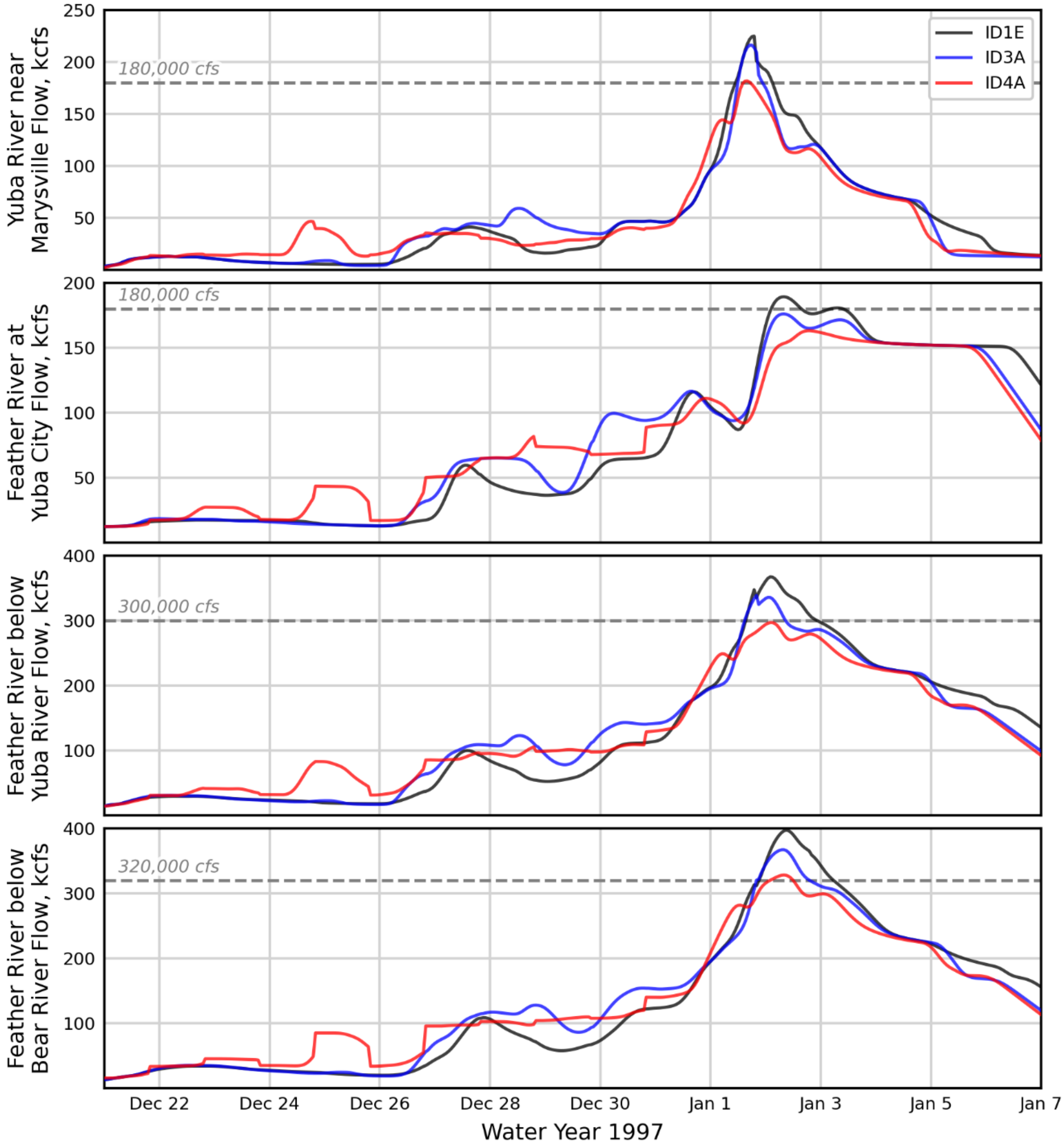
1997 108% Scaling
Oroville



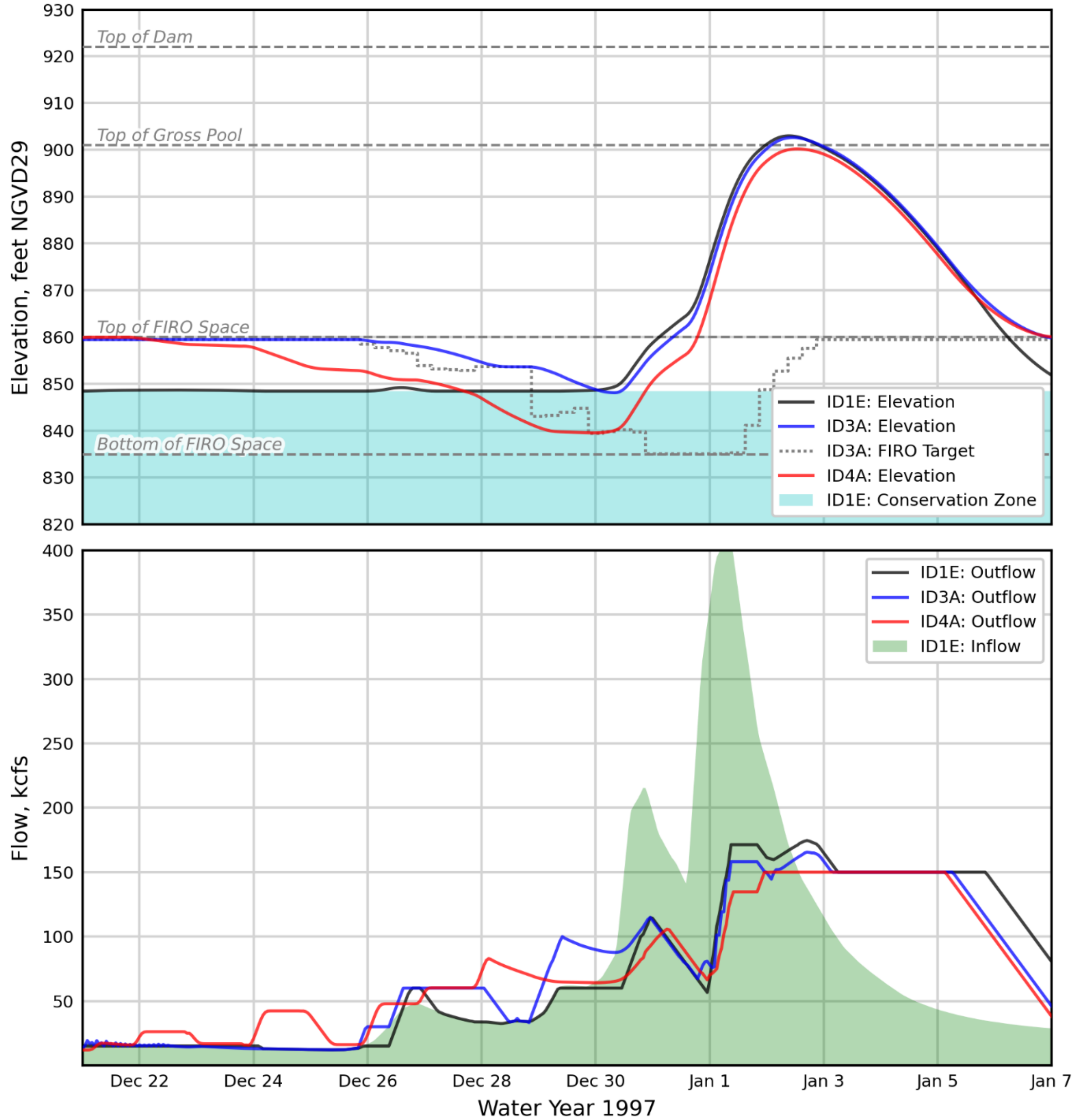
1997 108% Scaling
New Bullards Bar



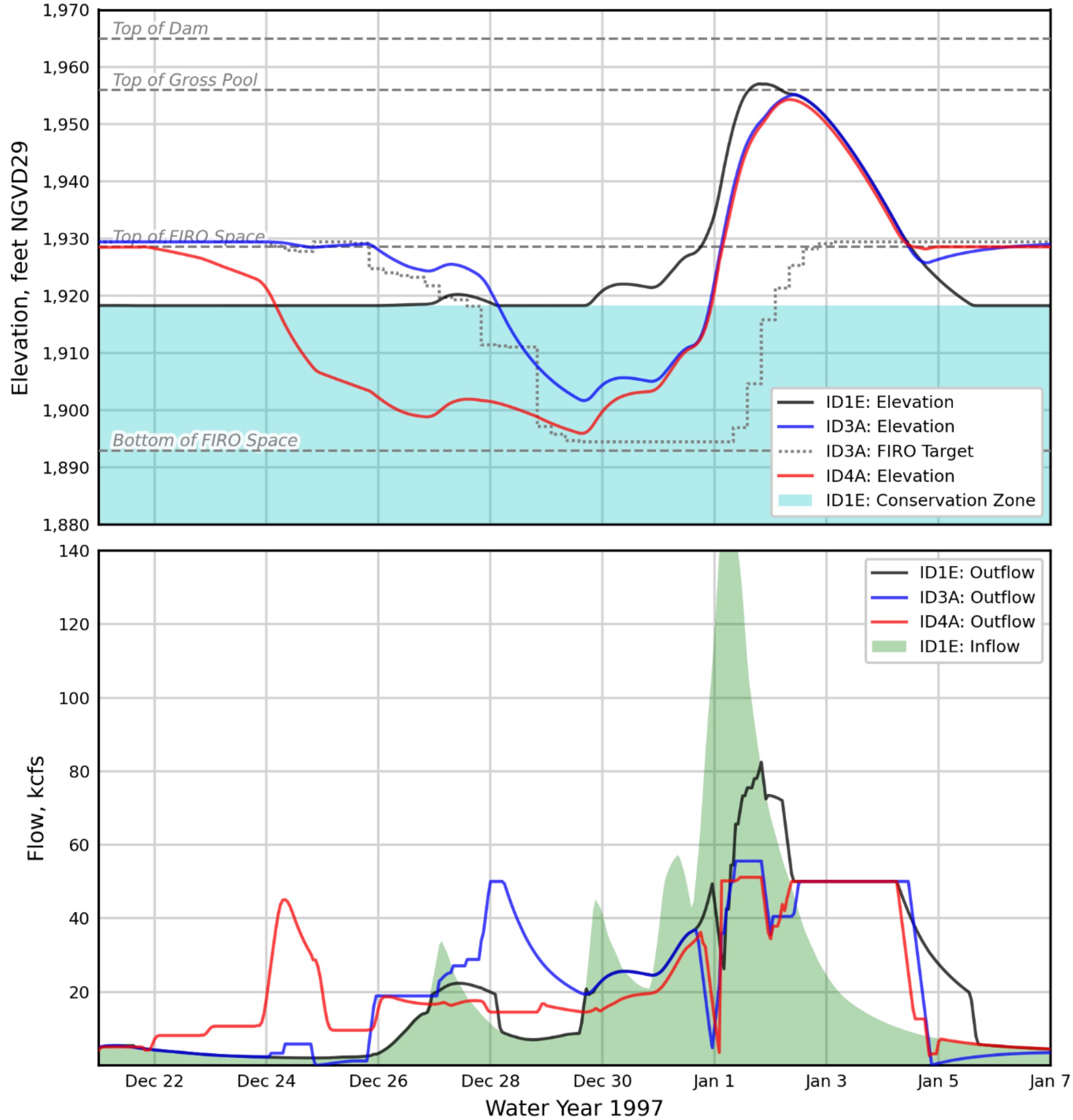
1997 110% Scaling
Downstream Control Points



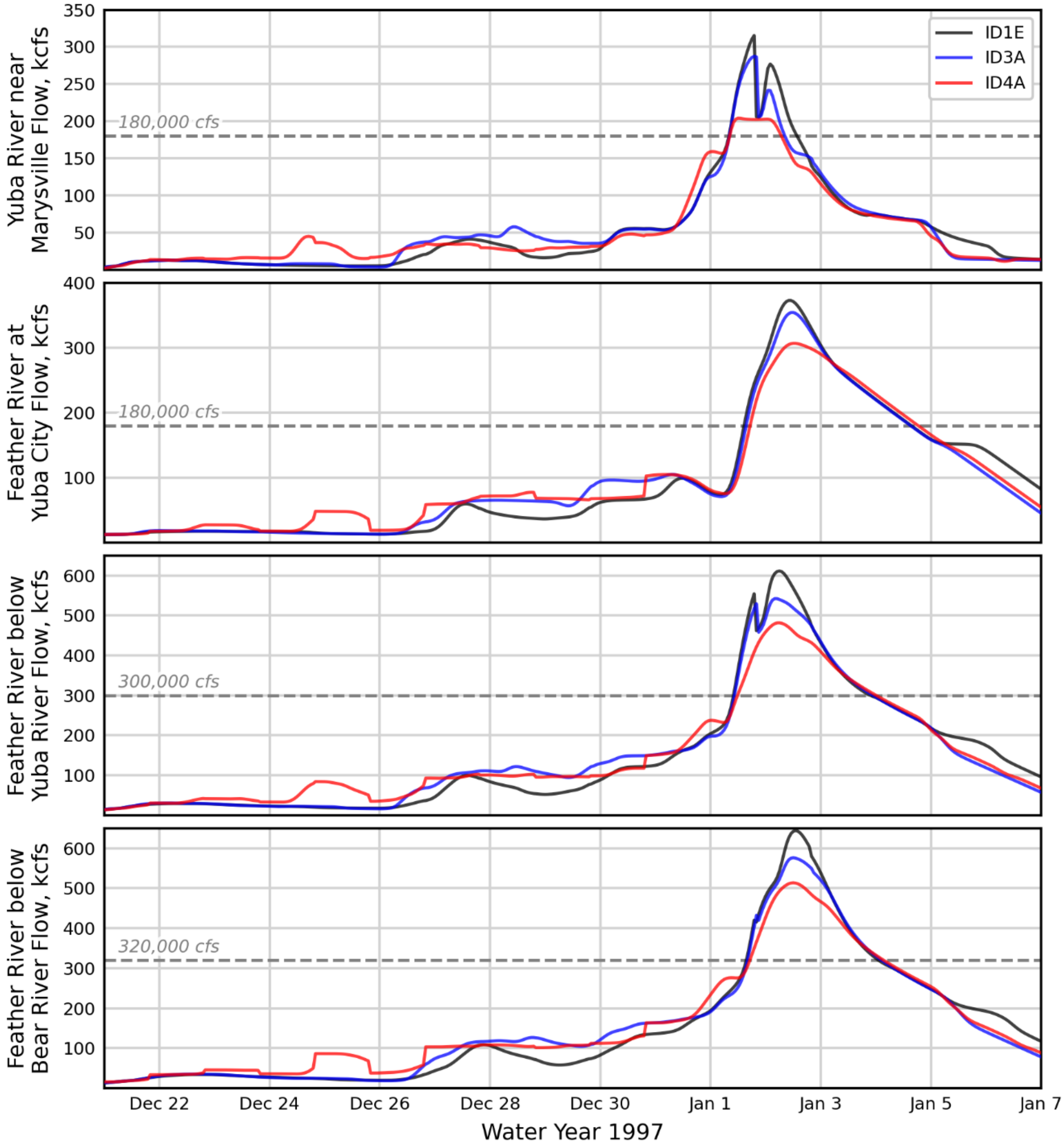
1997 110% Scaling
Oroville



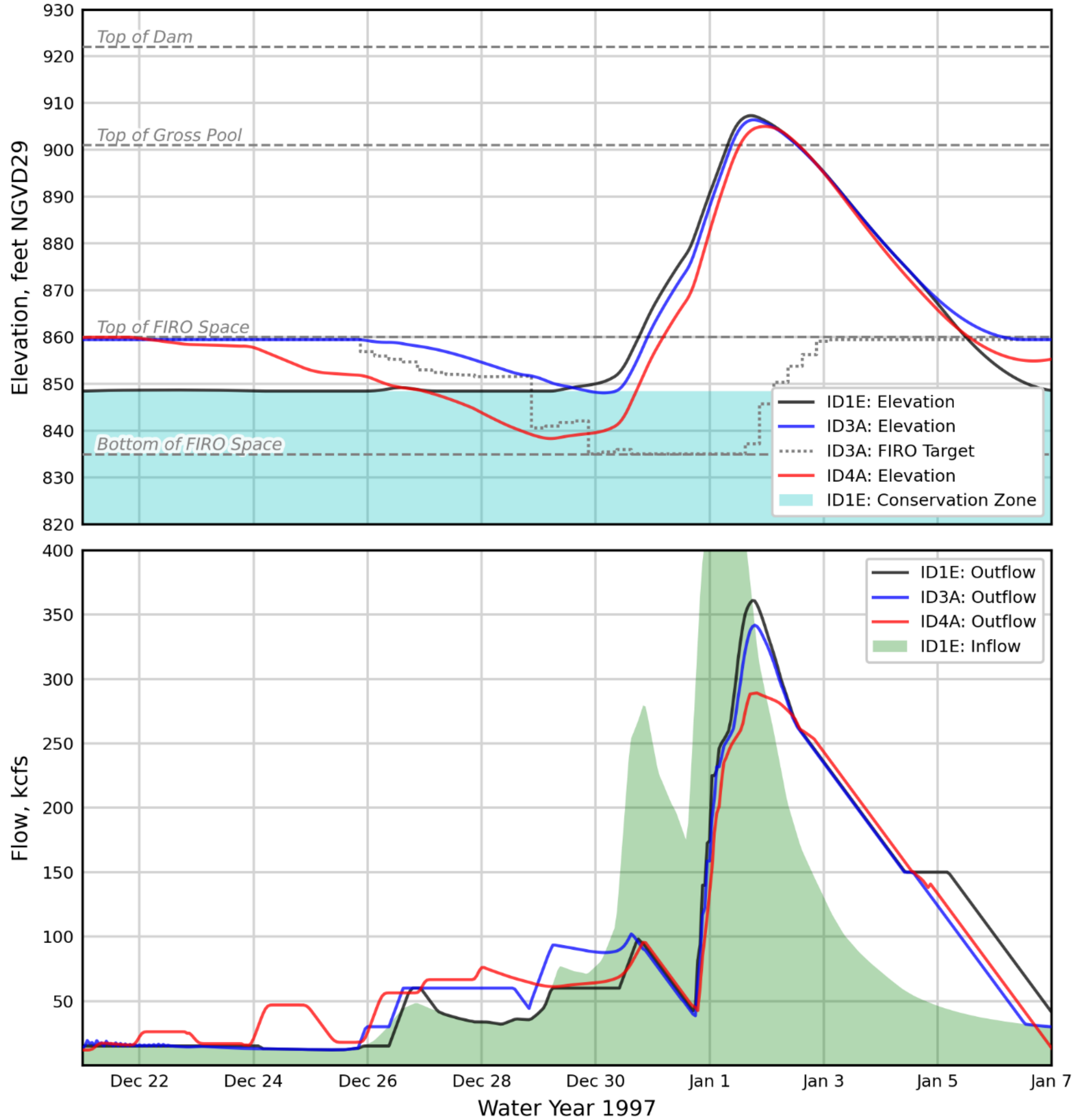
1997 110% Scaling
New Bullards Bar



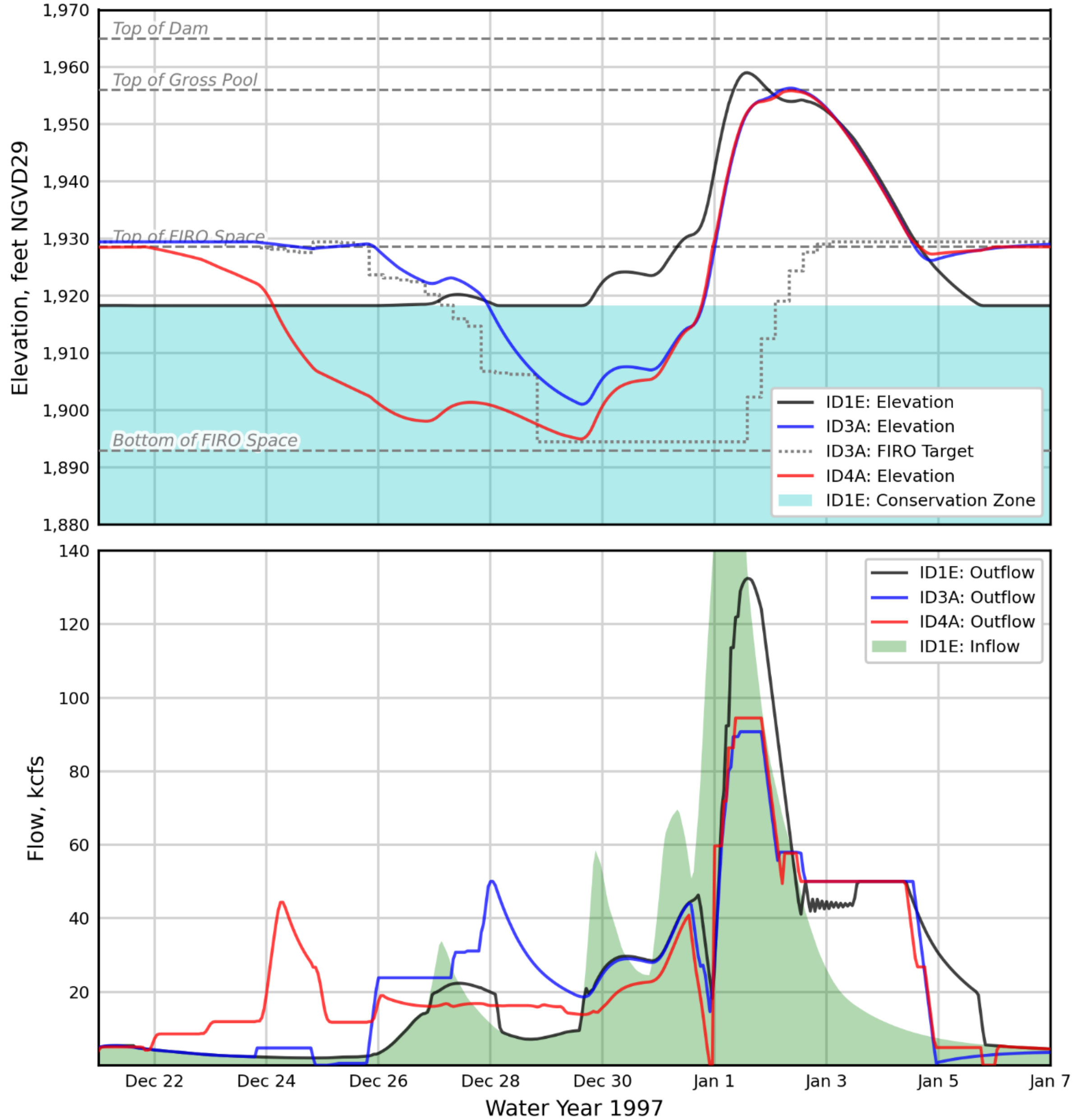
1997 120% Scaling
Downstream Control Points



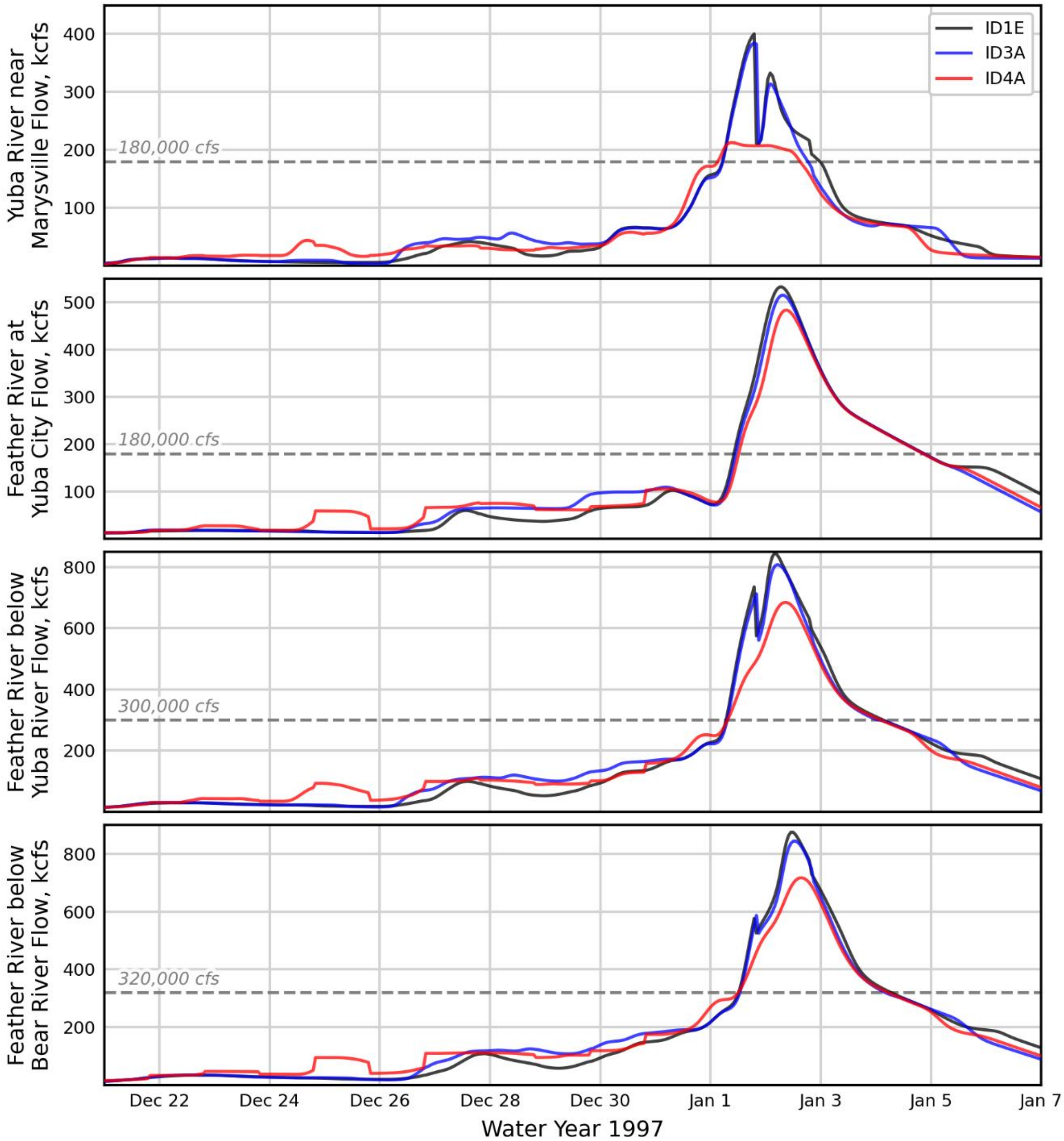
1997 120% Scaling
Oroville



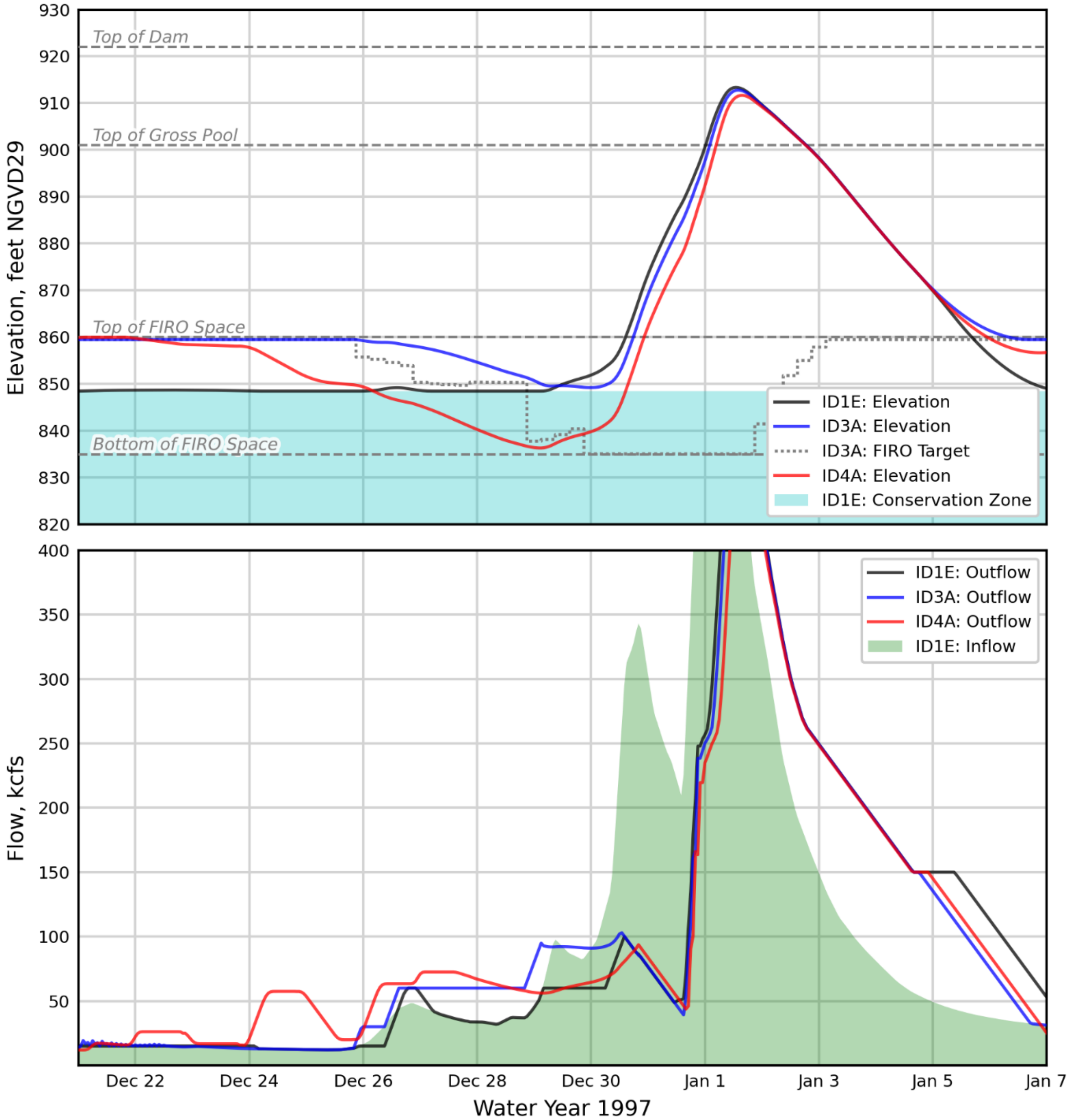
1997 120% Scaling
New Bullards Bar



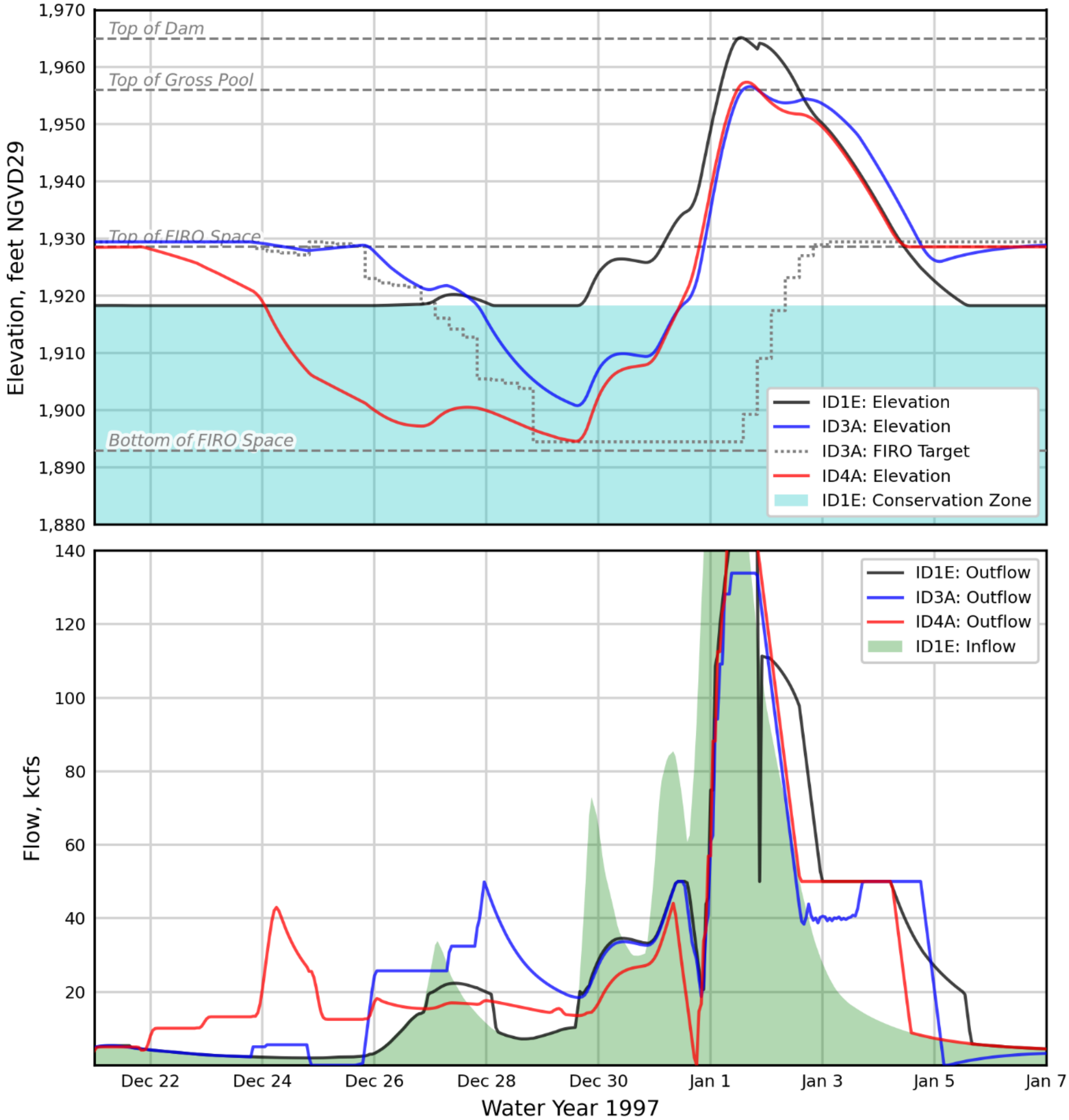
1997 130% Scaling
Downstream Control Points



1997 130% Scaling
Oroville



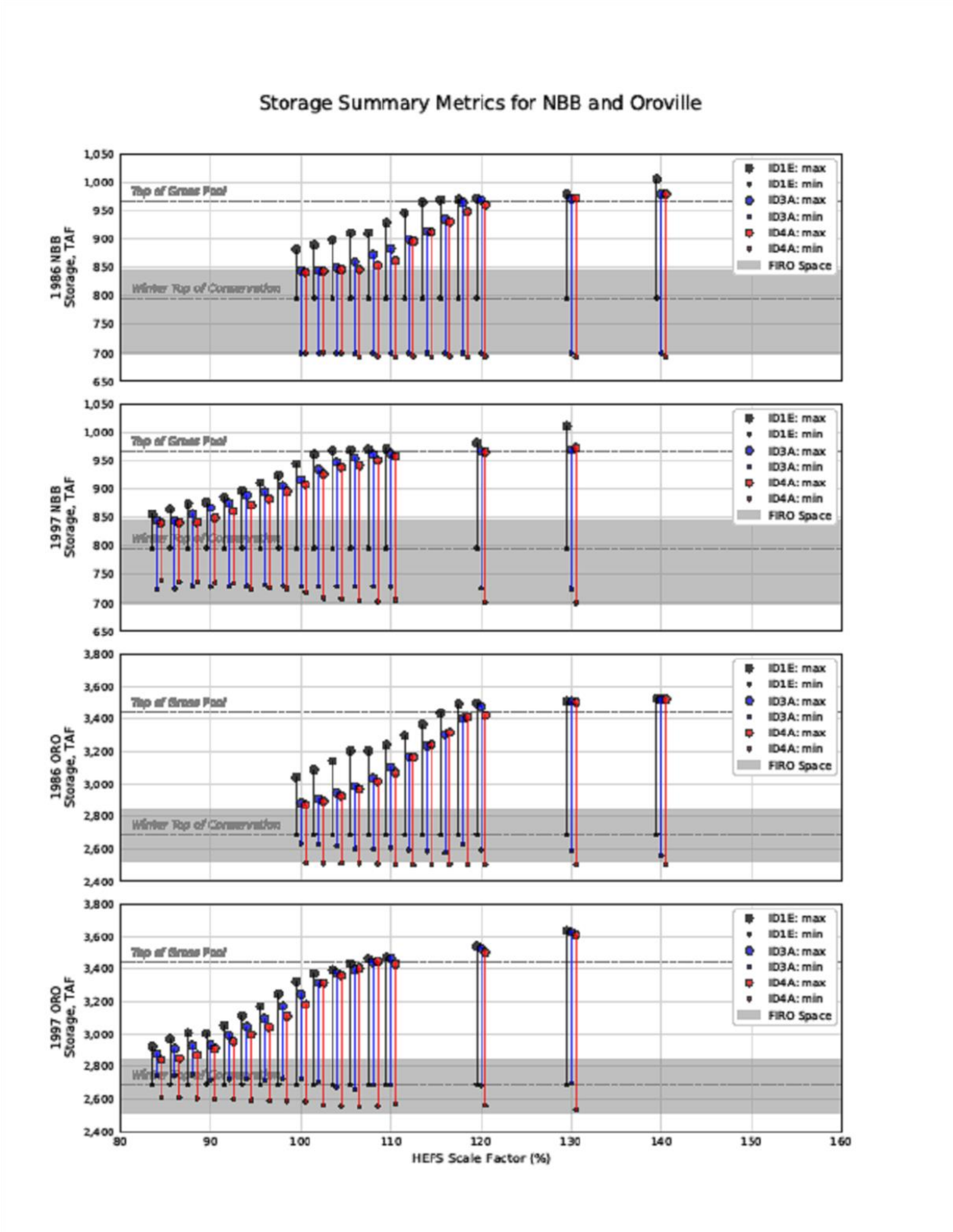
1997 130% Scaling
New Bullards Bar

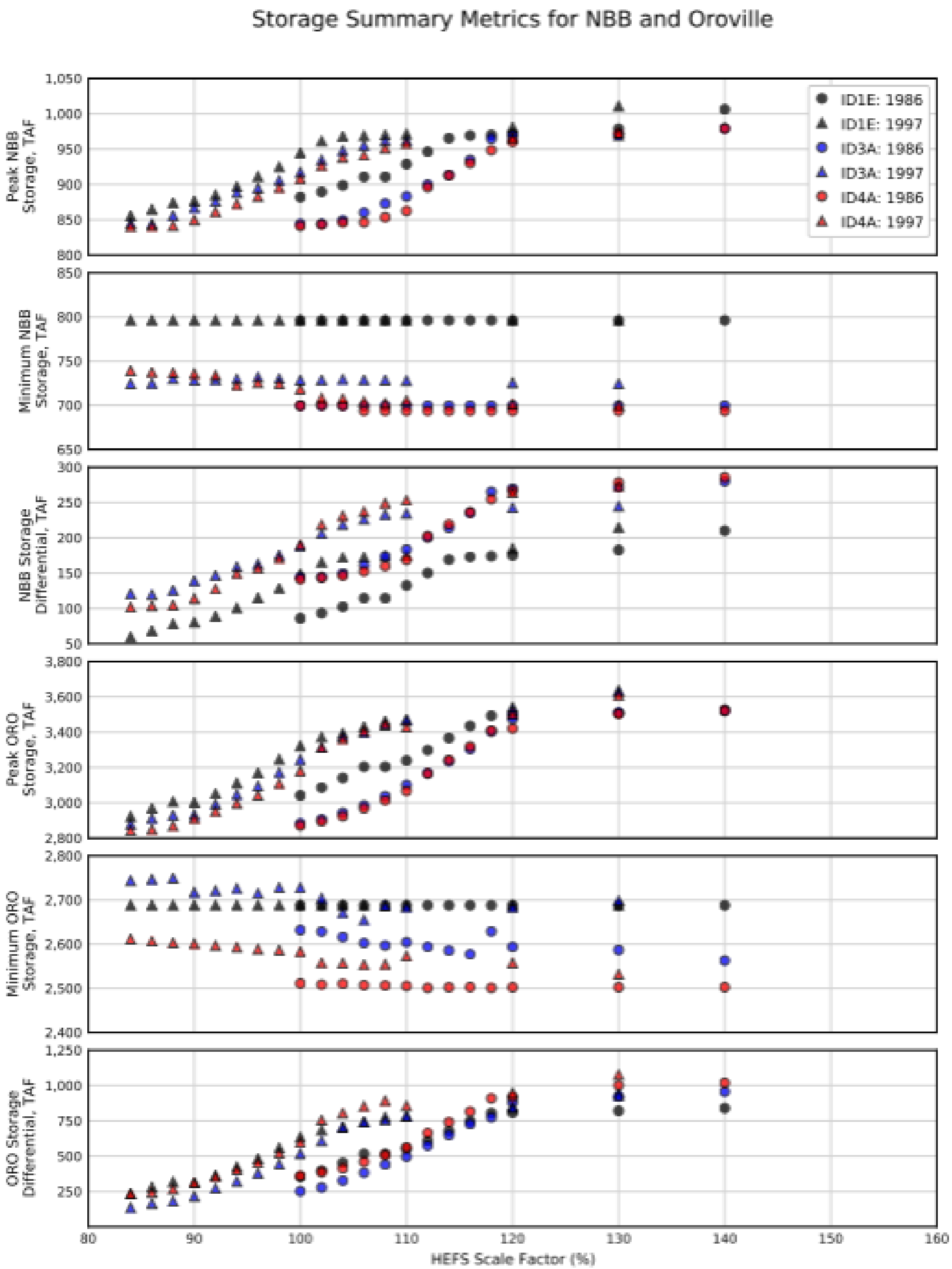


A.6 Storage Summary Metrics for Oroville and New Bullards Bar

1986 scale factors applied: 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 130, 140

1997 scale factors applied: 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 120, 130





A.7 Benefits Transfer Graphics

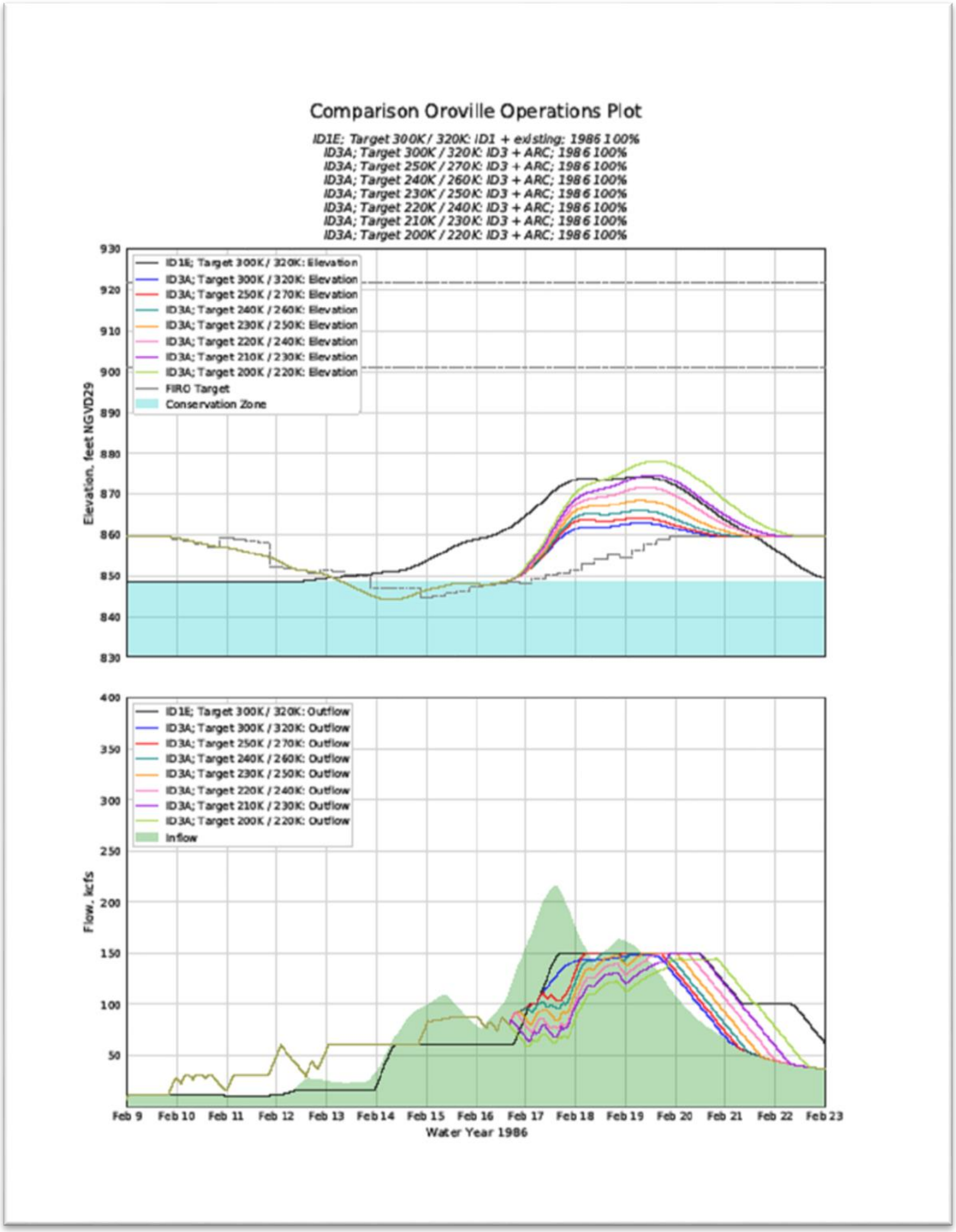
Process described in Section 4.3

Benefits transfer from reservoir storage to downstream control points demonstrated by reducing the target downstream flows in HEC-ResSim.

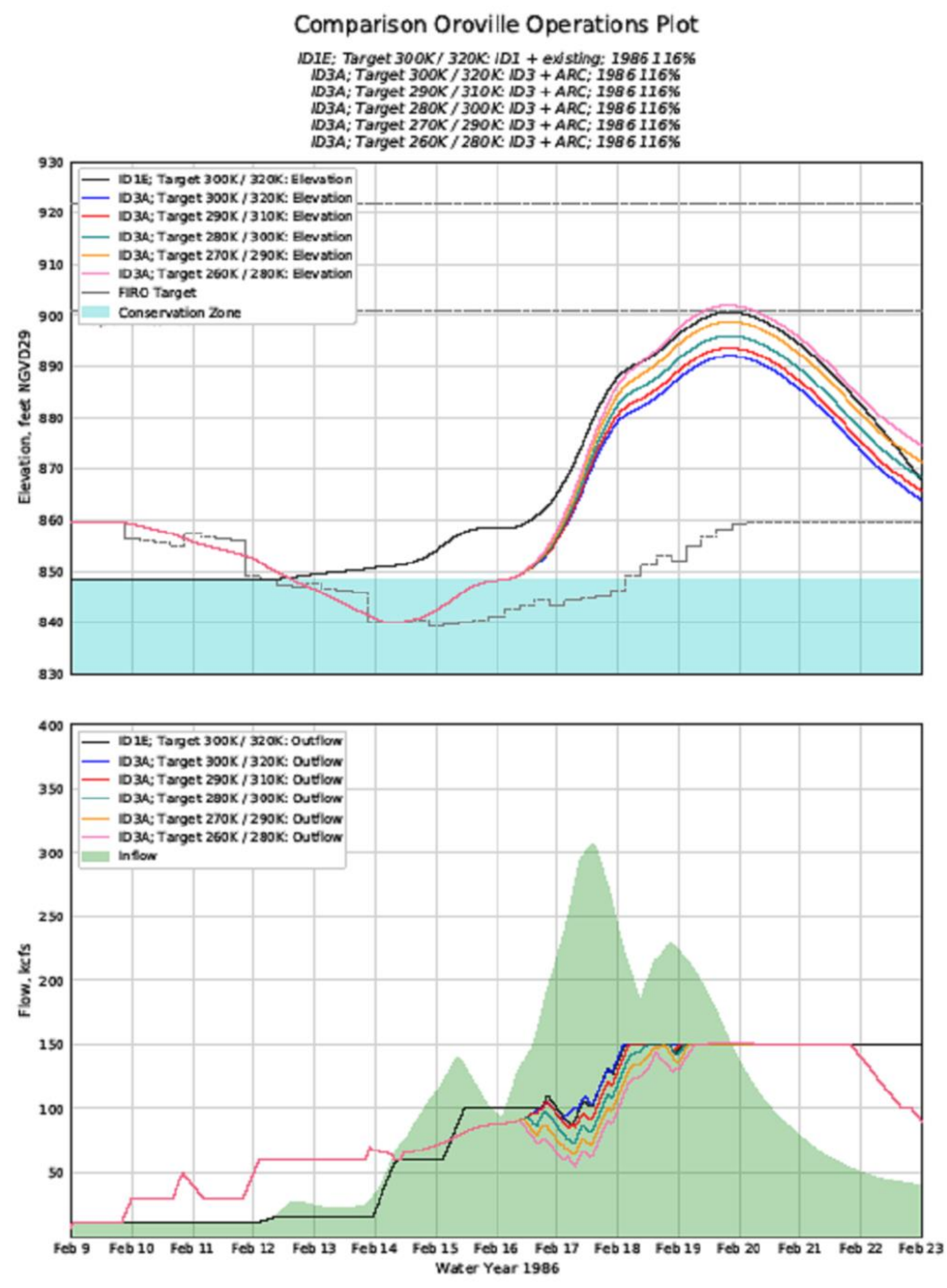
Only ID3A alternative was evaluated and compared with baseline (ID1E).

Graphic provided for 1986 scaled to 100% and 116% and 1997 scaled to 100% and 106%.

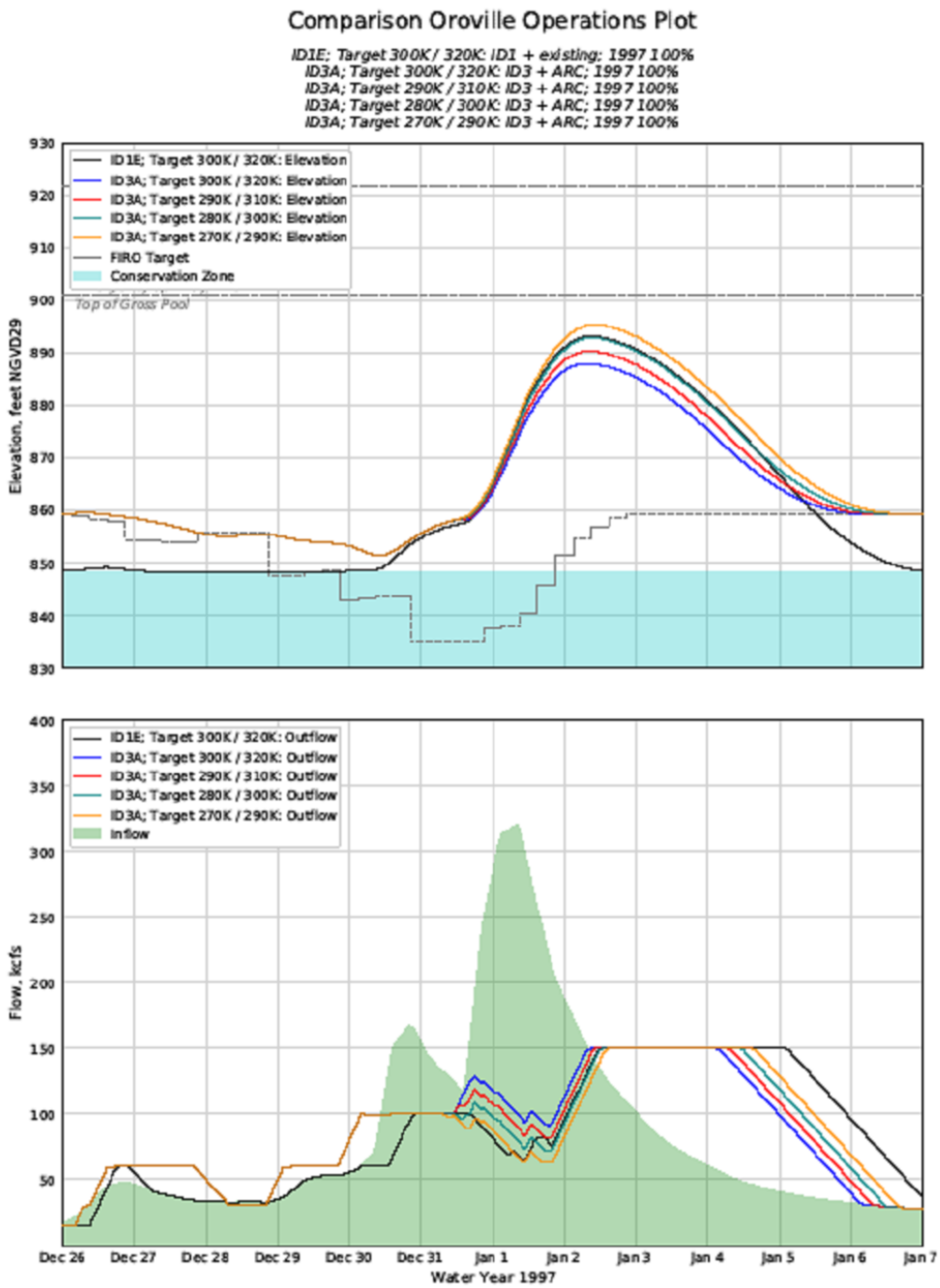
Benefits Transfer
100% Scaling of 1986 Event



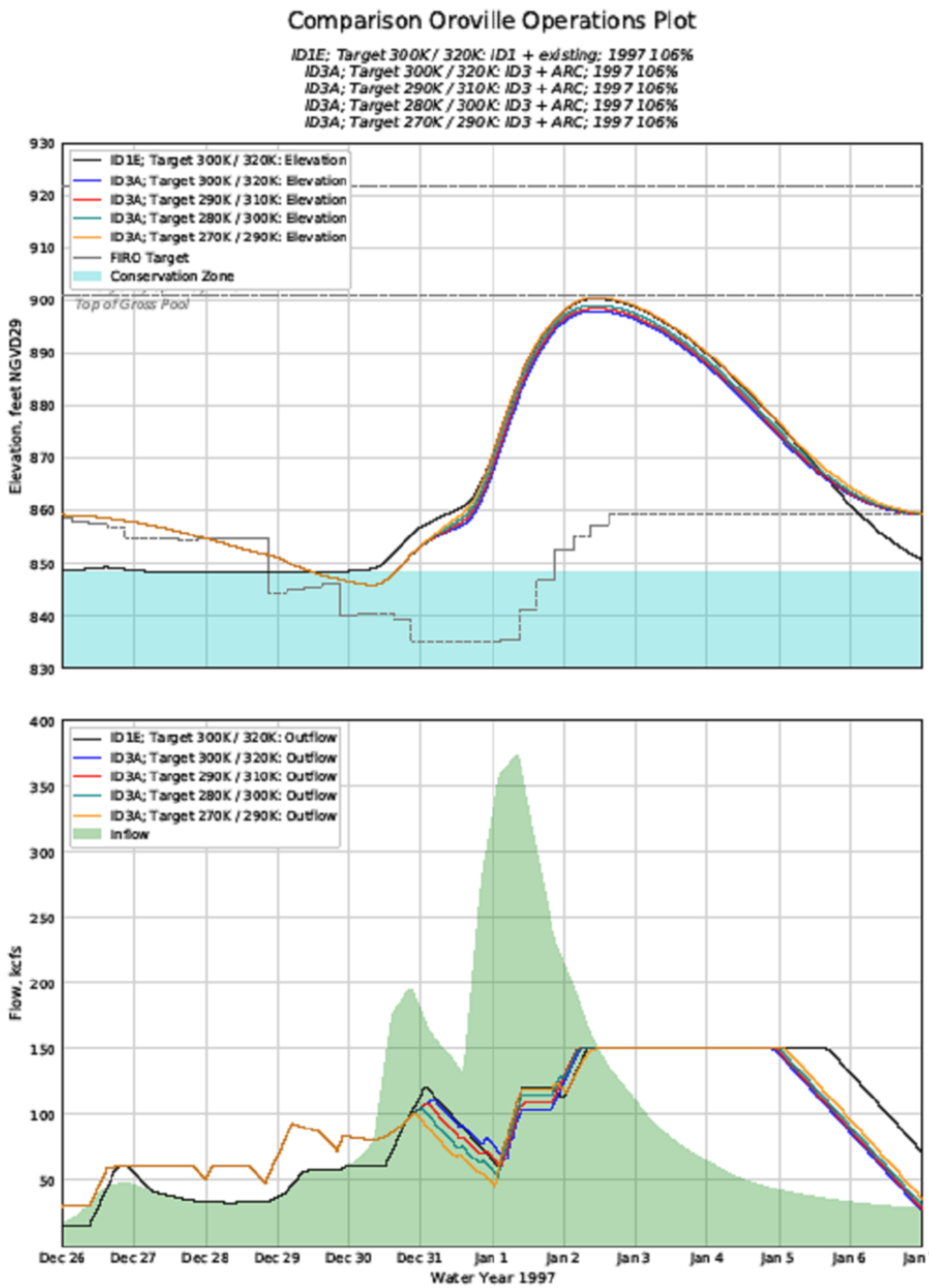
Benefits Transfer
116% Scaling of 1986 Event



Benefits Transfer
100% Scaling of 1997 Event



Benefits Transfer
106% Scaling of 1997 Event



Comparison Oroville Operations Plot
ID1E; Target 300K / 320K: ID1 + existing; 1997 106%
ID3A; Target 300K / 320K: ID3 + ARC; 1997 106%
ID3A; Target 290K / 310K: ID3 + ARC; 1997 106%
ID3A; Target 280K / 300K: ID3 + ARC; 1997 106%
ID3A; Target 270K / 290K: ID3 + ARC; 1997 106%
ID1E; Target 300K / 320K: Elevation
ID3A; Target 300K / 320K: Elevation
ID3A; Target 290K / 310K: Elevation
ID3A; Target 280K / 300K: Elevation
ID3A; Target 270K / 290K: Elevation
FIRO Target
Conservation Zone
Top of Gross Pool
Elevation, feet NGVD29
ID1E; Target 300K / 320K: Outflow
ID3A; Target 300K / 320K: Outflow
ID3A; Target 290K / 310K: Outflow
ID3A; Target 280K / 300K: Outflow
ID3A; Target 270K / 290K: Outflow
Inflow
Flow, kcfs
Dec 26
Dec 27
Dec 28
Dec 29
Dec 30
Dec 31
Jan 1
Jan 2
Jan 3
Jan 4
Jan 5
Jan 6
Jan 7
Water Year 1997

Appendix B – Meteorology

B.1 Precipitation and AR Catalog

Contributing Author: Chad Hecht, Center for Western Weather and Water Extremes

An extended (~20-year) catalog of precipitation, atmospheric river characteristics, and forecast verification statistics over the Yuba and Feather River watersheds was developed in a 72-hour and event accumulation perspective to further the understanding of the mechanisms that lead to precipitation and how it is forecast. The two forms of the catalog were developed utilizing mean-areal Stage-IV precipitation observations within the HUC-8 boundaries of the Yuba River and Feather River (combining the North Fork, East Branch North Fork, & Middle Fork) watersheds as the foundation of the catalog while several meteorological and atmospheric river related observations (derived from ERA5 Reanalysis and other observations) are provided, such as:

- Daily Mean IVT Magnitude and Direction
- Daily Maximum IVT Magnitude and Direction
- Atmospheric River Scale
- Time-integrated IVT and Direction
- Sierra-barrier Jet
- Freezing Level

Forecast information and statistics are included in conjunction with the observations listed above to contextualize and identify systematic sources of error as a function of lead time and physical process. Mean-areal quantitative precipitation forecasts from the West-WRF and GEFS were calculated for lead-times up to 48 hours identifying days and lead times that exhibited the largest and smallest forecast errors. In addition to precipitation forecasts, several variables and statistics were calculated from West-WRF data for atmospheric river characteristics, such as daily mean and maximum IVT magnitude and direction. The Method for Object-Based Diagnostic Evaluation (MODE) tool was utilized to identify landfall position error of atmospheric rivers and the role these errors played on the forecast of precipitation.

In summary, this overarching catalog serves as a meteorological reference for the Yuba and Feather River watersheds and the phenomena that can lead to extreme precipitation, over/under forecasts, short lead times, etc. The data within this catalog was utilized to perform several of the analyses presented in this FVA and will continue to provide information for studies and analyses in the future.

Event Rank	Event Start	Event End	Mean Areal Precipitation	Even Duration	Maximum 6 hour Precipitation	Time of Event Maximum Precipitation	Lower Watershed Precipitation	Upper Watershed Precipitation	Maximum IVT at the Coast	Time-Integrated IVT	AR Scale Coast
1	2017010712	2017011206	351.6476	114	40.248542	2017010900	284.46	477.87	1040.23	19.06	3
2	2017020600	2017021018	346.3108	114	[illegible]	[illegible]	195.42	426.03	390.62	20.15	4
3	[illegible]	[illegible]	[illegible]	108	[illegible]	[illegible]	246.86	355.09	557.86	17.57	3
4	2012031306	2012031900	321.3971	144	[illegible]	[illegible]	205.25	328.99	471.34	11.72	2
5	2002121312	2002121806	281.8541	114	[illegible]	[illegible]	170.17	302.48	937.63	16.45	4
6	2014020606	2014021018	271.7008	108	42.77248	2014020900	206.75	290.81	641.3	12.14	3
7	2021102306	2021102606	265.1605	72	[illegible]	2021102418	195.34	[illegible]	1169.55	10.10	4
8	2019021306	2019021800	231.8581	114	35.111949	2019021412	135.18	301.87	878.64	12.79	4
9	2022122912	2023010100	230.3845	66	[illegible]	2022123100	180.50	287.12	506.12	10.53	3
10	2005123006	2006010100	225.5238	42	[illegible]	[illegible]	136.56	345.65	983.55	8.46	3

Table B-1 An example of the information provided within the Yuba Watershed Event Catalog showing the top 10 events from 2002 to 2023.

B.2. Evaluation of physical and mesoscale processes

B.2.1 Heavy precipitation days in the Upper Yuba are driven by ARs

Contributing Author: Paul Loikith, Portland State University

Activity: We have characterized storm types and atmospheric patterns associated with and those leading up to heavy precipitation days in the Upper Yuba Watershed. Results improve our understanding of the atmospheric drivers of heavy precipitation with implications for prediction in the watershed.

Summary of Results: Storm types/atmospheric patterns associated with the five days leading to heavy precipitation days in the Upper Yuba Watershed are grouped into clusters using the self-organizing map (SOM) method. Figure 1 (left) shows a nine-node SOM of integrated water vapor transport (IVT) for all five-day periods ending in a day where the basin-wide mean daily accumulated precipitation was above the 90thpercentile of all >=2mm precipitation days (i.e., heavy precipitation days) between the months of October and March spanning the years 1980-2022. To construct the SOM, the daily mean IVT for each heavy precipitation day along with the preceding four days was provided as input to the algorithm which then assigned each pentad of days to one of the nine nodes (or clusters) such that days with a similar progression of IVT characteristics are grouped together. The IVT maps in the nine panels of Figure 1 are approximately the composite mean of all the IVT maps for all days assigned to each node. For more background on implementing the SOM method in this way see Loikith et al. (2017) and Aragon et al. (2020).

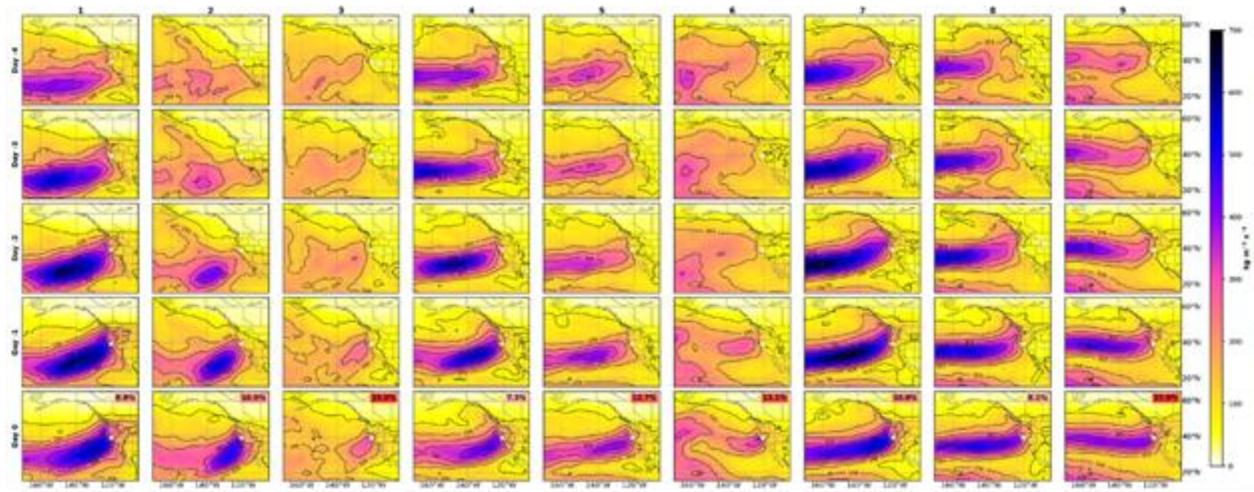


Figure B-1*. (Left) A 9-node SOM of IVT for the five days leading up to a heavy precipitation day in the Upper Yuba Watershed. The nodes are referred to by the number above each row. The number at the top of each panel in the bottom row is the percent of all extreme precipitation days assigned to that node. The bottom row shows the IVT pattern for the day recording extreme precipitation which each row above shows the IVT pattern for the days prior. The SOM algorithm clusters the entire pentads such that days assigned to Node 1 have a 5-day progression like the one shown in the left column.*

All nine nodes depict atmospheric river (AR)-like features with narrow corridors of elevated IVT directed at the Upper Yuba Watershed. However, there is a considerable range of IVT strength, orientation, and length of the IVT corridors as well as the progression of the pattern in the days prior. For example, days assigned to Node 1 (left) are characterized by very high IVT values (relative to the other nodes) with the corridor of enhanced IVT oriented from southwest to northeast on the extreme day (bottom row). This AR-like feature strengthens and extends eastward in the days leading to the extreme. This can be understood to be capturing strong ARs with subtropical moisture origin. On the other hand, days assigned to Node 3 are characterized by weaker IVT with the corridor of enhanced transport following a cyclonic pattern without an obvious direct subtropical connection. The enhanced IVT develops near the coast during the day prior to the extreme.

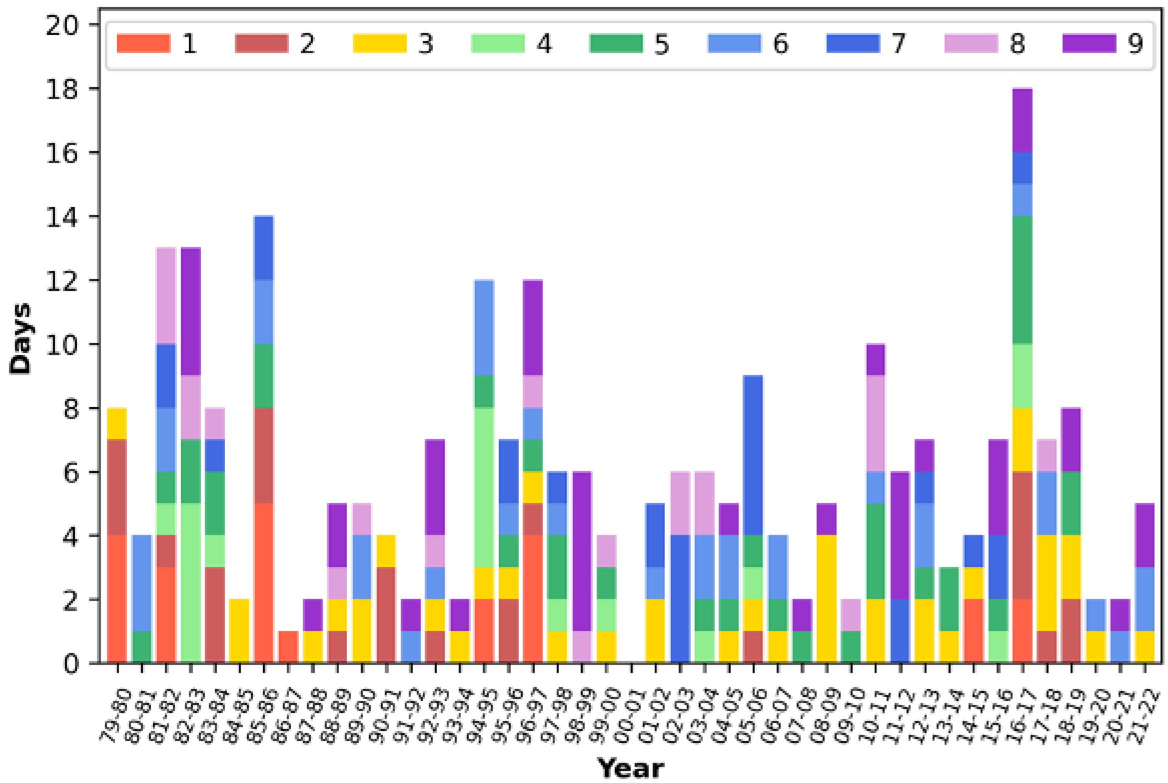


***Figure B-2**. The number of days assigned to each node per water year.*

Figure B-2 shows the observed distribution of node assignments for heavy precipitation days by water year. All years had at least one day that exceeded the 90th percentile of daily precipitation (1984 and 2013 only had one day). For years with multiple heavy precipitation days, there is a range of node assignments, although some years show preference towards IVT pattern type. For example, nearly half of 2017 heavy precipitation days were assigned to nodes 2 and 5. It is worth noting that 2017 had the greatest number of heavy precipitation days with notable hydrological impacts within the region.

With the IVT-based SOM constructed (Figure B-1), other variables that may help diagnose heavy precipitation can be composited for days assigned to each node's 5-day progression. Figure B-3 shows composites for 300 hPa wind, and sea level pressure (SLP). 300 hPa wind patterns help diagnose the upper-level support for heavy precipitation and AR behavior. For example, Node 2 shows a jet streak rounding the bottom of a trough, coinciding with surface cyclogenesis in the corresponding SLP maps where upper-level divergence would be expected. This further diagnoses the development of elevated IVT in the day prior to the extreme day. For each node, the 5-day sequence of synoptic conditions associated with ITV-trained SOM provided a physical diagnosis of the conditions leading to the extreme precipitation day.

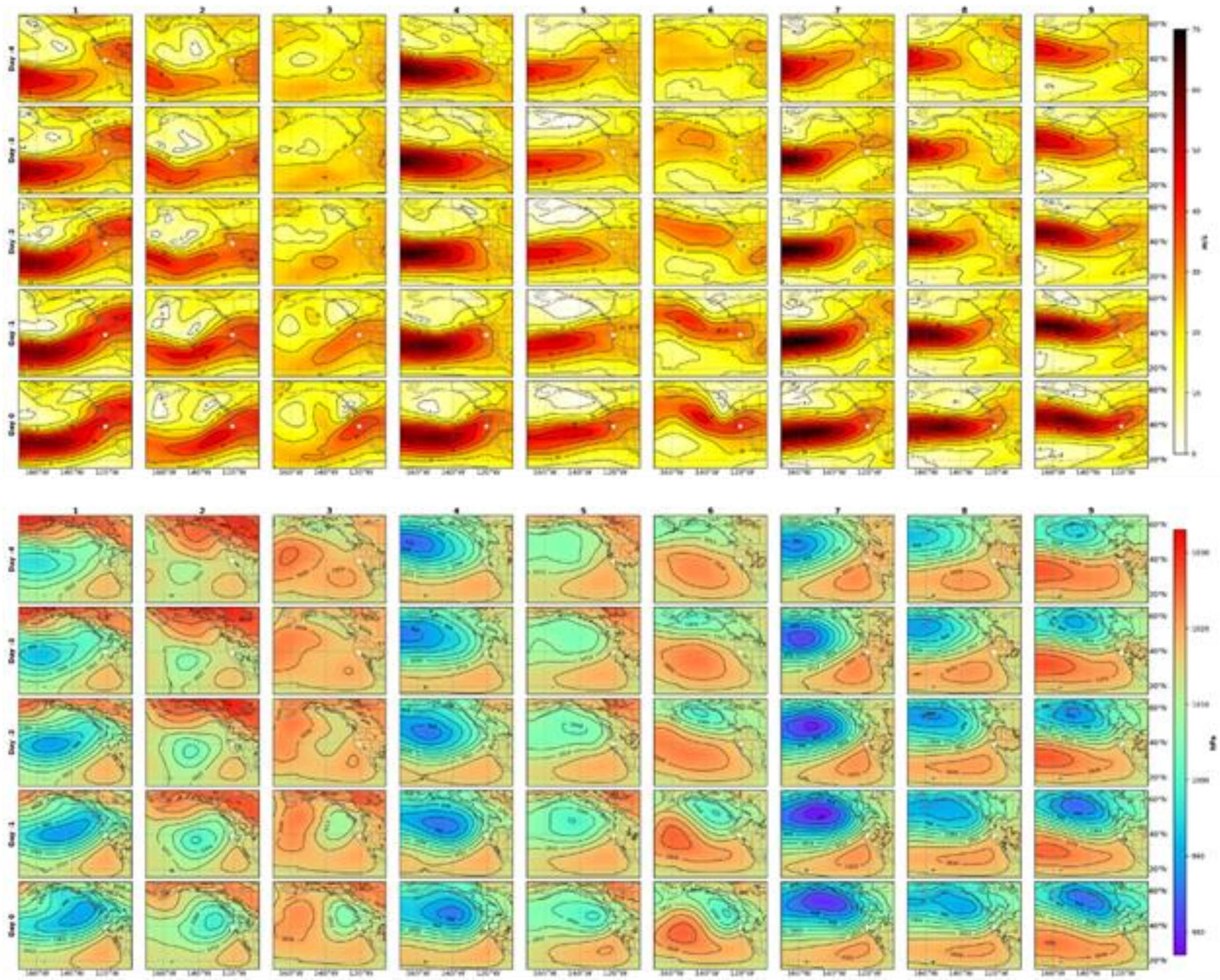


Figure B-3. *Composite mean of (top) 300 hPa wind speed and (bottom) sea level pressure for the days assigned to each node.,*

Outcomes

- Paper in preparation: Russell, E., and P. C. Loikith, 2023: Synoptic drivers of heavy precipitation days in the Upper Yuba Watershed of California. *In preparation for Journal of Hydrometeorology.*
- This research activity motivates assessment of the utility of the observations-based SOM in predicting heavy precipitation events within the watershed using large-scale patterns as predictors.
- This research activity motivates expanding this methodology to other watersheds as the range of storm types and characteristics will likely differ based on physical geography and local climatology.

References

Aragon, C., P. C. Loikith, N. McCullar, and A. Mandilag, 2020: Large Scale Meteorological Patterns Associated with Extreme Precipitation Events over Portland, OR. *Int. J. Climatol.*, **40**, 4763-4780.

Loikith, P. C., B. Lintner, and A. Sweeney, 2017: Characterizing Large-Scale Meteorological Patterns and Associated Temperature and Precipitation Extremes over the Northwestern United States using Self-Organizing Maps. *J. Climate*, **30**, 2829-2847.

B.2. Evaluation of physical and mesoscale processes

B.2.2 Upslope water vapor flux along ARs drives variability in precipitation

B.2.3 Precipitation generally increases in elevation in the Upper Yuba

B.2.4 Precipitation varies spatially across the Feather sub-basins

B.2.5 Sierra Barrier Jet is responsible for a portion of the Feather sub-basin precipitation

Note: B.2.2, B.2.3, B.2.4, B.2.5 are combined here

Contributing Authors: Chris Castallano, Center for Western Weather and Water Extremes and Chad Hecht, Center for Western Weather and Water Extremes

The following analysis will examine how topography and storm characteristics influence the spatial variability of precipitation in the Yuba-Feather system. The spatial distribution of precipitation in the Upper Yuba subbasin is evaluated by dividing the subbasin into four elevation bands (based on the lower, middle, and upper quartiles of elevation), and calculating the observed precipitation in each elevation band. Elevations corresponding to the three quartiles are estimated from the 4-km PRISM (Daily et al. 2008) digital elevation model. As is typically observed in regions of complex terrain, precipitation generally increases with elevation in the Upper Yuba. During the 20 water years (WYs) spanning 2003-2022, the average total WY mean areal precipitation (MAP) in the lower and upper portions of the Upper Yuba was 882 mm (34.7 inches) and 1552 mm (61.1 inches), respectively (Figure B-4).

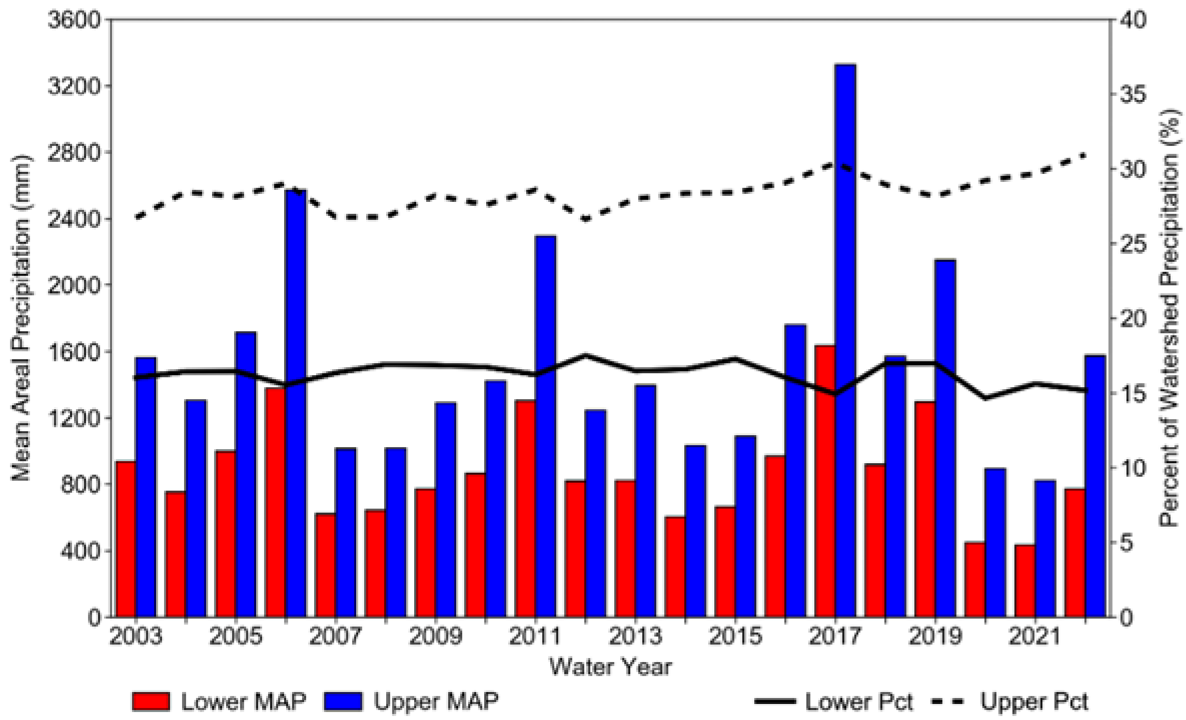


Figure B-4: Time series showing the total water year mean areal precipitation (MAP) in the lower (red bars) and upper (blue bars) portions of the Upper Yuba subbasin. Lines represent the relative contribution from the lower (dotted) and upper (solid) portions to the total water year precipitation in the Upper Yuba.

Overall, about 29% of the total precipitation fell in the upper 25% of the Upper Yuba, whereas only 16% of the total precipitation fell in the lower 25% of the Upper Yuba. The relative contributions from the upper and lower portions of the Upper Yuba to the total WY precipitation are anti-correlated (i.e., WYs with greater contribution from the lower portion of the Upper Yuba are characterized by less contribution from the upper portion of the Upper Yuba, and vice versa), with a correlation coefficient of −0.67.

As Figure B-5 illustrates, the relative contributions from the upper and lower portions of the Upper Yuba to total event precipitation are also anti-correlated. The strength of the anti-correlation increases with precipitation event magnitude. For example, the correlation coefficient for all events is −0.76, but the correlation coefficient for events with more than 5 inches of MAP in the Upper Yuba is −0.86.

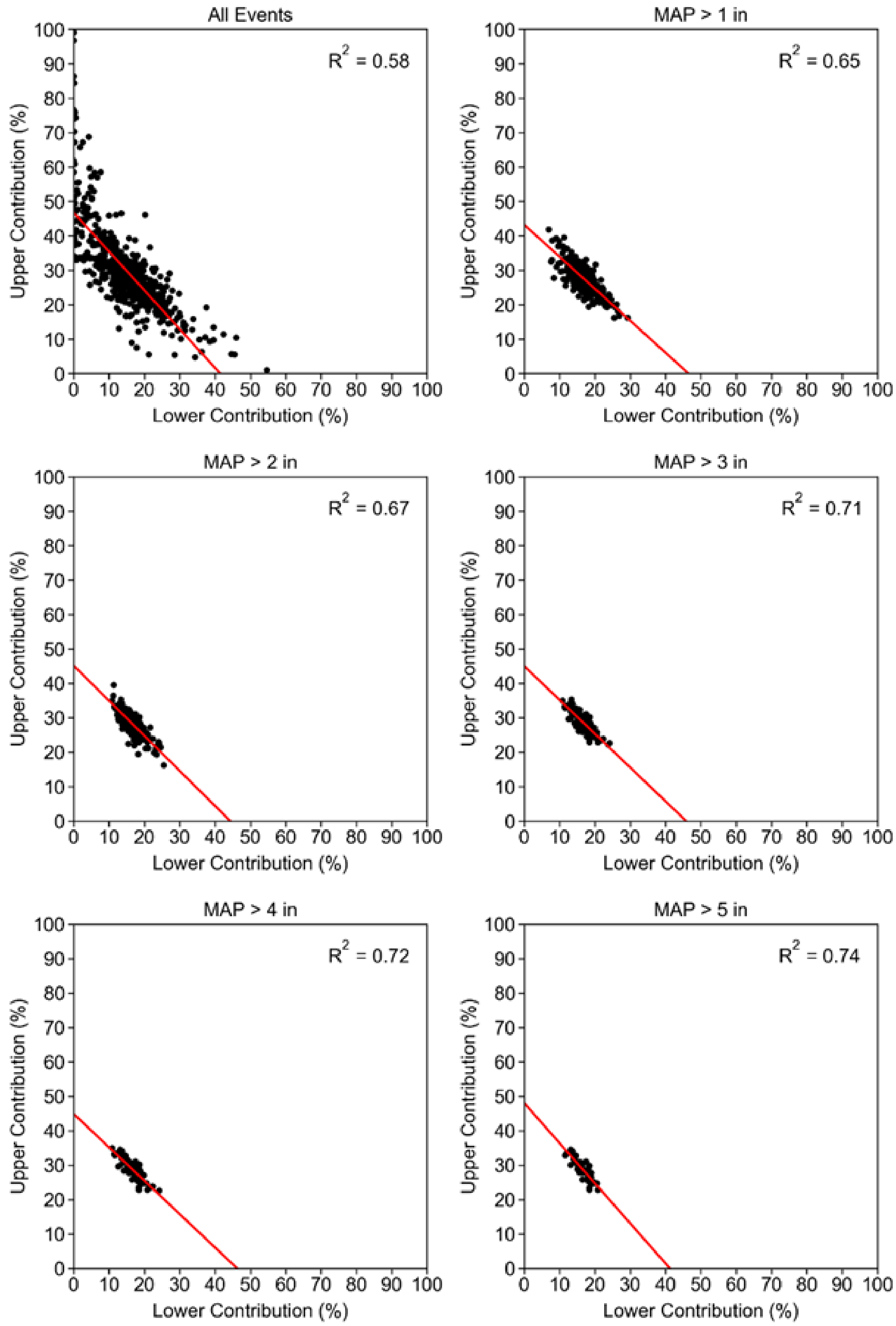


Figure B-5: Scatterplots showing the relative contributions to total event precipitation from the lower (x-axis) and upper (y-axis) portions of the Upper Yuba subbasin. Linear regression lines are plotted in red, and the square of the correlation coefficient is provided in the upper right corner of each plot.

The spatial distribution of precipitation in the Feather watershed is evaluated by calculating the observed precipitation in each of the three Feather subbasins - the North Fork Feather (NFF), the East Branch North Fork Feather (EBNFF), and the Middle Fork Feather (MFF). As Figure B-6 demonstrates, the Feather watershed's precipitation climatology varies considerably by subbasin. Climatologically, the NFF is the wettest subbasin, followed by the MFF. The EBNFF is substantially drier than the other two subbasins due to being located on the leeward side of the main spine of the Sierra Nevada, where rain shadowing is common.

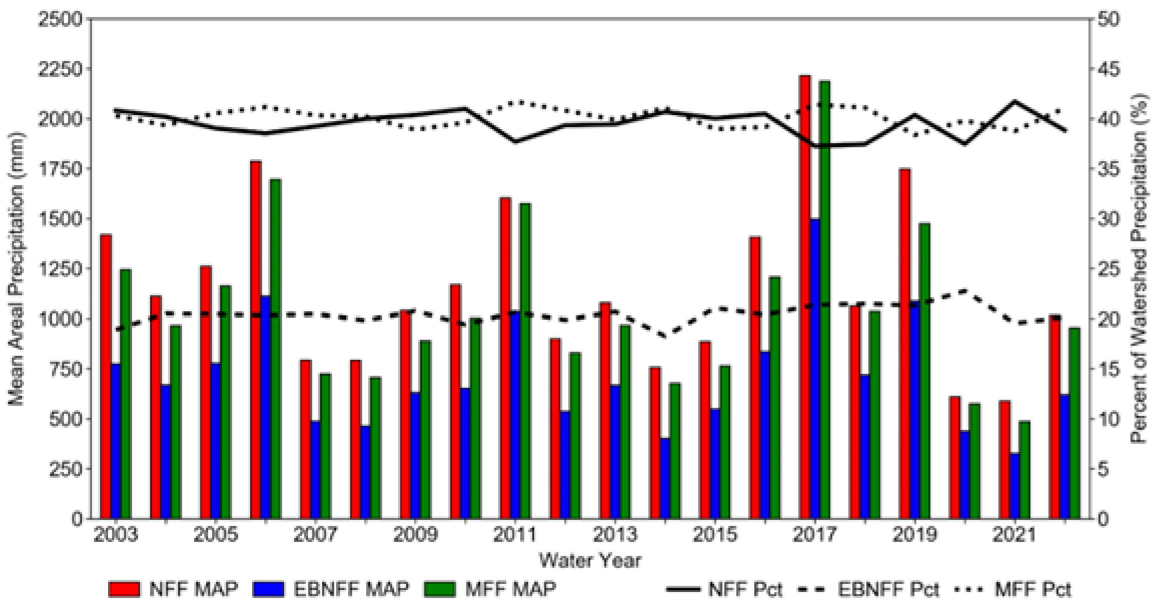


Figure B-6: Time series showing the total WY mean areal precipitation (MAP) in the North Fork Feather (red bars), East Branch North Fork Feather (blue bars), and Middle Fork Feather (green bars). Lines represent the relative contribution from the North Fork Feather (solid), East Branch North Fork Feather (dashed), and Middle Fork Feather (dotted) to the total WY precipitation in the Feather system.

During the 20 water years (WYs) spanning 2003-2022, the average total WY mean areal precipitation (MAP) in the NFF, MFF, and EBNFF subbasins was 1162 mm (45.7 inches), 1056 mm (41.6 inches), and 714 mm (28.1 inches), respectively. Overall, 39% of the total precipitation fell in the NFF, 40% of the total precipitation fell in the MFF, and only 20% of the total precipitation fell in the EBNFF. Note that the relative contribution from the MFF is slightly higher than the relative contribution from the NFF because the MFF is larger in area.

The relative contribution from the NFF to the total WY precipitation is anti-correlated (r = −0.66) with the relative contribution from both the MFF and the EBNFF to the total WY precipitation. As Figure B-5 illustrates, the relative contributions from the NFF and MFF to total event precipitation are also anti-correlated, and the strength of the anti-correlation remains steady for different event precipitation thresholds. A different relationship is observed between the NFF and EBNFF, such that the strength of the anti-correlation is much weaker for events with more than 2 inches of MAP in the Feather watershed. These results suggest that there are certain storm characteristics that determine whether the heaviest precipitation falls in the NFF or the MFF.

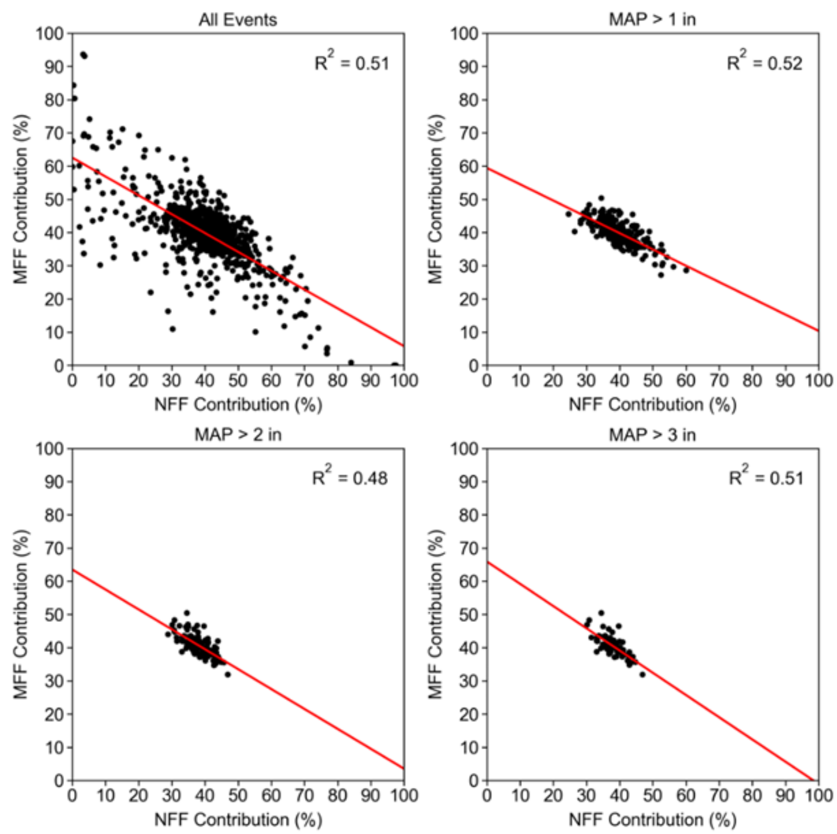


Figure B-5: Scatterplots illustrating the relative contributions to total Feather event precipitation from the North Fork Feather (x-axis) and Middle Fork Feather (y-axis). Linear regression lines are plotted in red, and the square of the correlation coefficient is provided in the upper right corner of each plot.

Numerous studies have demonstrated that heavy precipitation in Northern California is often associated with landfalling atmospheric rivers (ARs). Comparing the relationship between Mean-areal event precipitation and projected time-integrated IVT (TIVT) at the mouth of the Yuba River watershed shows that TIVT explains 91% (R^2=0.912) of the variability in storm total precipitation (Figure B-6). Additionally, we investigate the role that AR intensity may have on the distribution of precipitation across the Yuba River watershed. Figure B-7 shows how the relative contribution from each of the four elevation bands in the Upper Yuba varies based on the event maximum AR-related IVT at a grid point near San Francisco, CA.

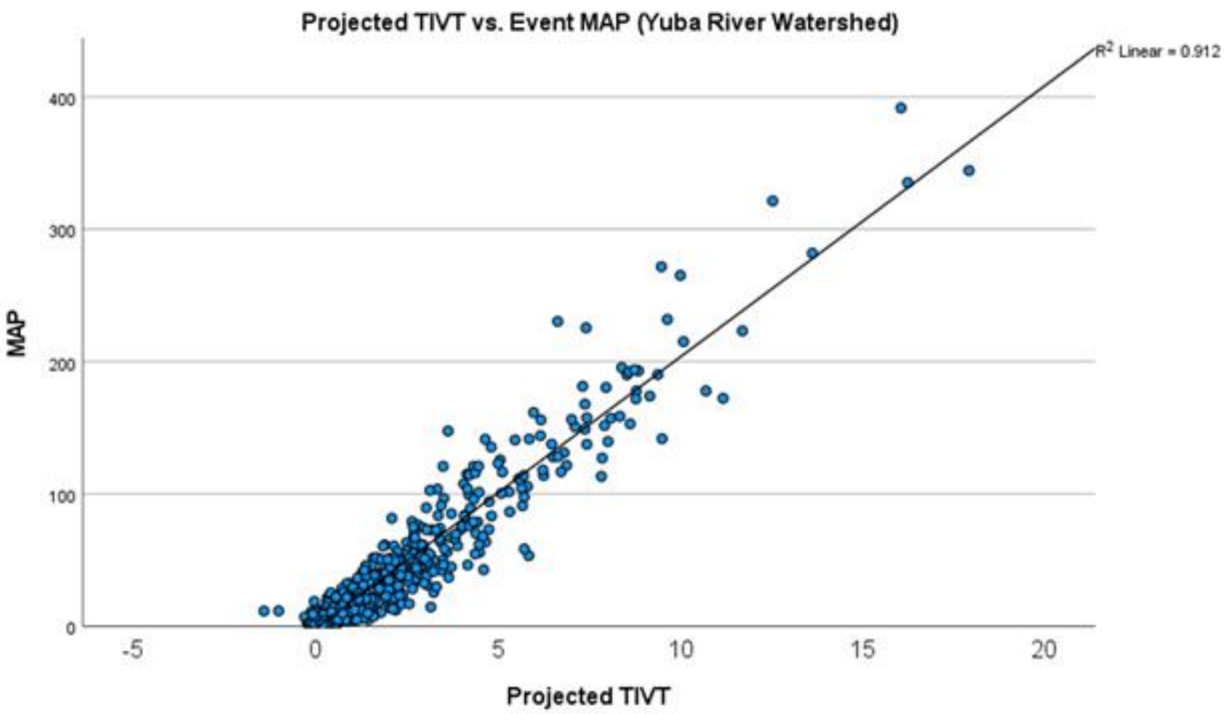


Figure B-6: Yuba River watershed event Mean-areal precipitation (mm) vs. projected time-integrated IVT (10^7 kg m^{-1}) onto 225 degrees (southwest) at the mouth of the Yuba River watershed.

For events with no AR (i.e., AR intensity and duration do not meet the minimum AR Scale criteria), the spread in the relative contributions from each elevation band is quite large. It is worth noting that these events are generally characterized by smaller MAP in the Upper Yuba and include some warm-season events with limited areal precipitation coverage. Compared to events with a weak or moderate-strength AR, events accompanied by a strong AR (maximum IVT ≥ 750 kg m^{-1} s^{-1}), are characterized by a small but statistically significant ($p < 0.05$ based on Mood's median test) increase (decrease) in the relative contribution from the upper (lower-to-middle) 25% of the subbasin.

Similar differences are observed for events that feature an AR3, AR4, or AR5, versus an AR1 or AR2 on the AR Scale. These results imply greater orographic enhancement of precipitation in the Upper Yuba during the strongest ARs, perhaps due to increased moisture flux at altitudes near the Sierra crest. Applying a similar analysis to the subbasins in the Feather watershed, we find a statistically significant increase (decrease) in the relative contribution from the EBNFF (NFF) during events featuring a strong AR versus events featuring a weak or moderate AR (Figure B-8). While the explanation for this finding is not yet clear, one possibility is that moisture transport associated with stronger ARs may be able to penetrate further inland, rather than being blocked by the higher terrain in the central parts of the NFF and MFF (see Figure B-9).

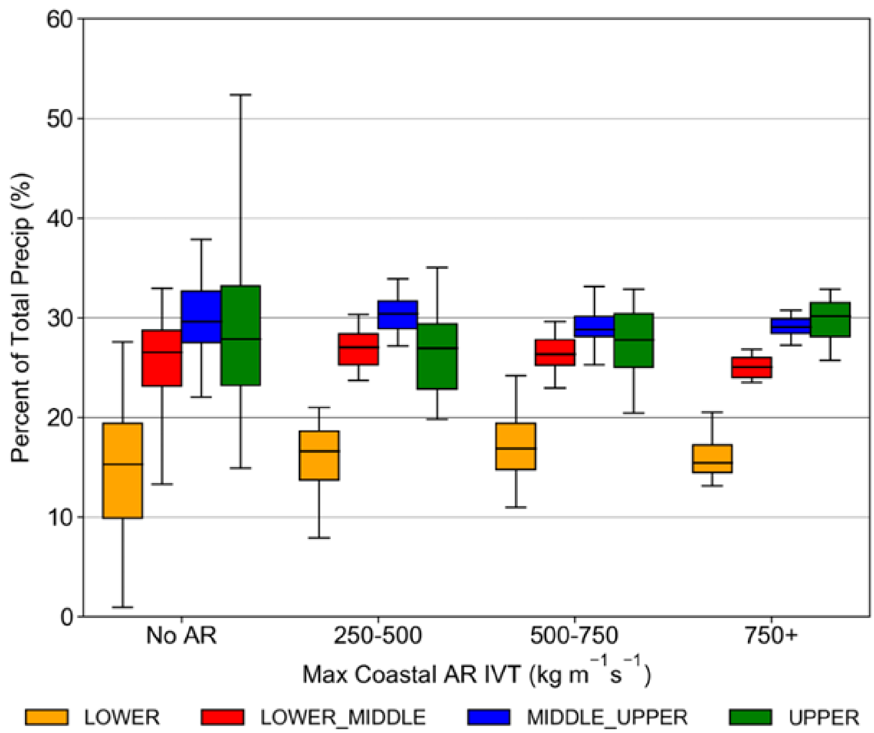


Figure B-7: Box plots showing the statistical distribution of the percent of total event precipitation falling in each of the four elevation bands in the Upper Yuba for different categories of AR intensity at 37.5°N, 122.5°W. Horizontal lines denote the median values. Boxes represent the interquartile range (IQR). Whiskers denote the lowest (highest) values above (below) the lower (upper) quartile minus (plus) 1.5 times the IQR.

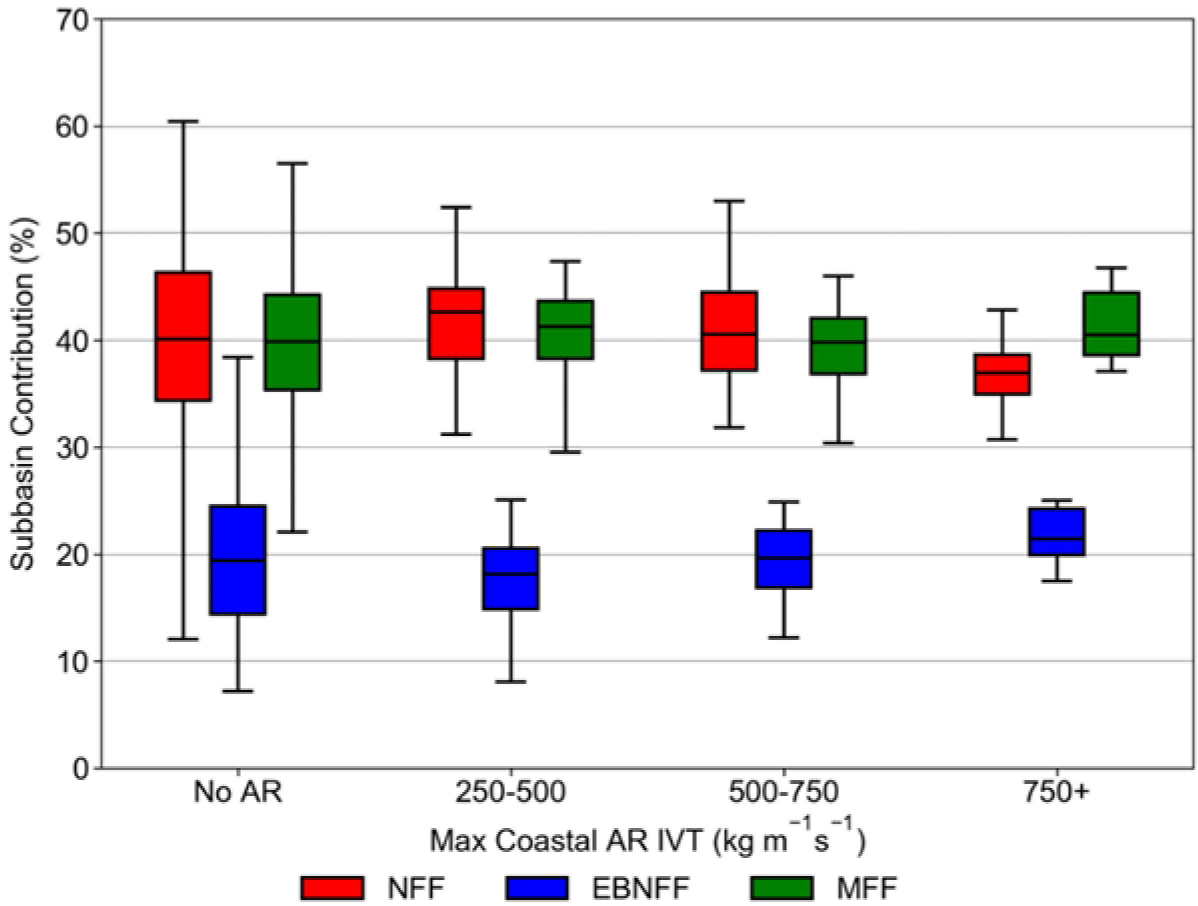


Figure B-8: Box plots showing the statistical distribution of the percent of total Feather event precipitation falling in each of the three Feather subbasins for different categories of AR intensity at 37.5°N, 122.5°W. Horizontal lines denote the median values. Boxes represent the interquartile range (IQR). Whiskers denote the lowest (highest) values above (below) the lower (upper) quartile minus (plus) 1.5 times the IQR.

Another meteorological phenomenon that can modulate precipitation in Northern California is the Sierra barrier jet (SBJ). Previous research by Neiman et al. (2013) found that SBJs, which are often observed with ARs penetrating through the Bay Area gap, can increase precipitation amounts at the northern end of the Sacramento Valley due to enhanced low-level southerly moisture transport.

Here, we investigate the sensitivity of the spatial distribution of precipitation within the Yuba-Feather system to SBJs by comparing event precipitation at two stations, one in the NFF (Four Trees), and another in the Upper Yuba (Alleghany). These stations have similar elevations and are both situated between the valley floor and the main spine of the Northern Sierra Nevada. Both stations are in locations favorable for orographic enhancement from synoptic-scale southwesterly moisture transport associated with landfalling ARs, but the nearly east-west orientation of the Northern Sierra Nevada as it crosses through the NFF suggests that low-level southerly moisture transport associated with SBJs may provide additional precipitation forcing at Four Trees. Therefore, we may expect enhancement of precipitation at Four Trees relative to precipitation at Alleghany during events featuring an SBJ. This hypothesis is generally supported by the results shown in Figure B-10. Compared to events without an SBJ, the slope of the linear regression is much steeper for events featuring both an AR and an SBJ. In other words, as event precipitation increases, the relative increase in precipitation at Four Trees is much larger than the relative increase in precipitation at Alleghany when both an AR and an SBJ are present. Interestingly, events with an SBJ but no AR do not exhibit the same behavior, which suggests that both an AR and an SBJ must be present to produce this precipitation effect. Finally, it is worth noting that the largest precipitation events at both stations predominantly occur when both an AR and an SBJ are present, and nearly all events featuring an AR3, AR4, or AR5 on the AR Scale also feature an SBJ.

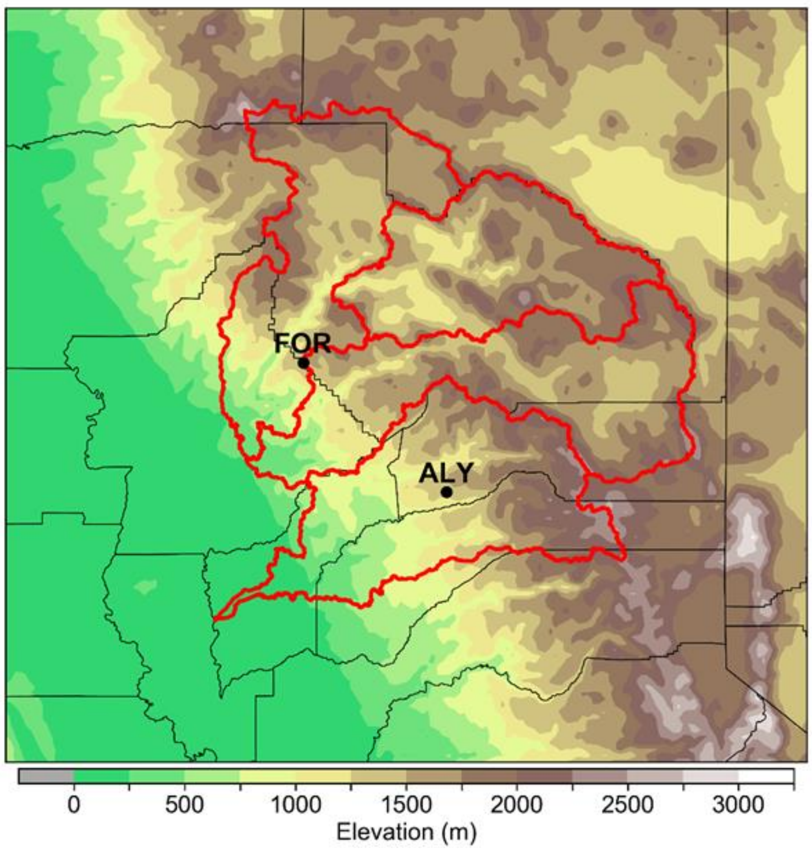


Figure B-9: Map showing elevation (color shading), county boundaries (black polygons), and the outlines of the HUC8 subbasins in the Yuba-Feather system (red polygons). Black circles denote the locations of the Four Trees (FOR) and Alleghany (ALY) precipitation gauges.

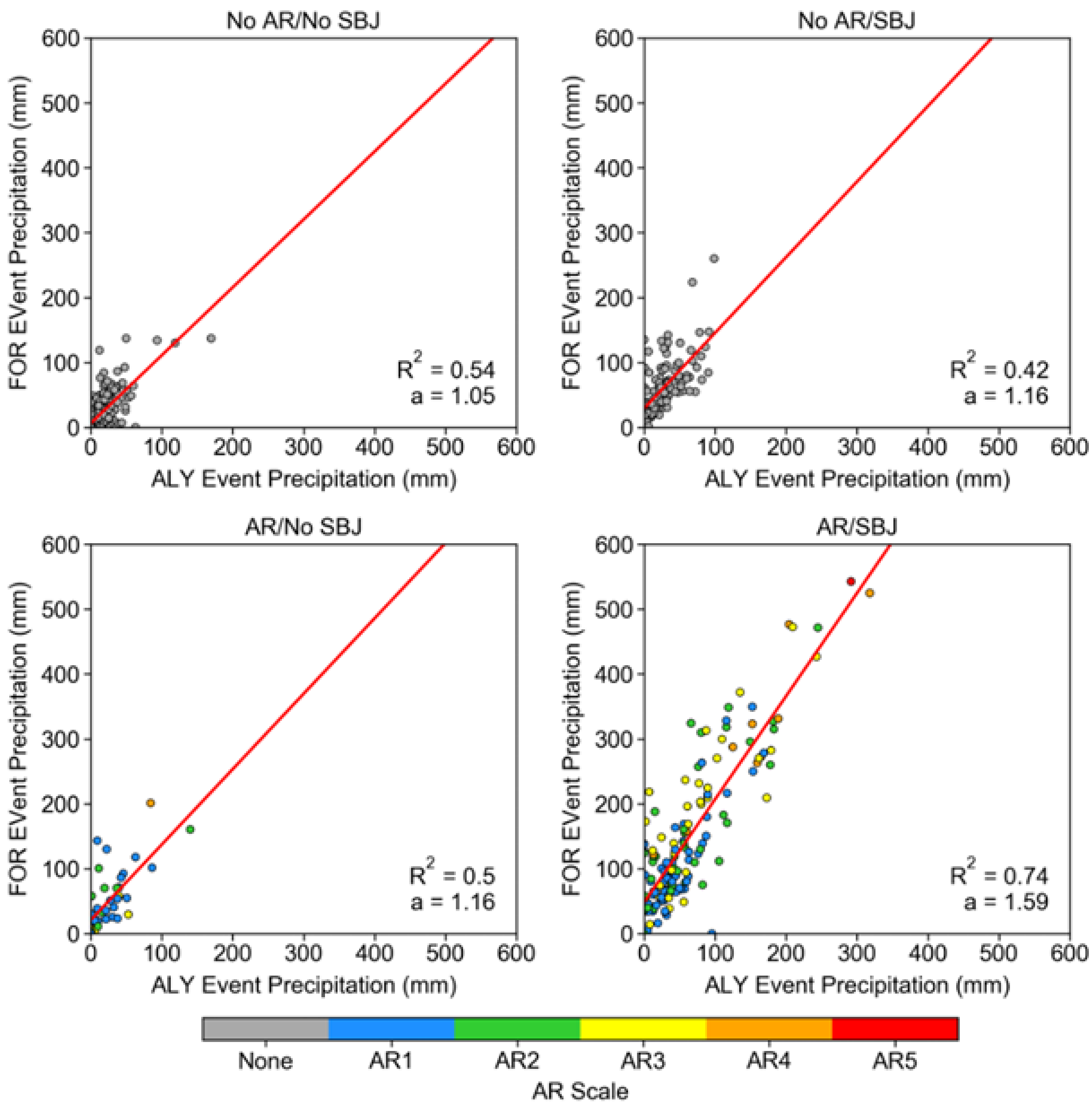


Figure B-10: Scatterplots showing the total precipitation observed at the Alleghany (ALY; x-axis) and Four Trees (FOR; y-axis) stations during precipitation events in the Feather watershed featuring: no AR and no SBJ (top left), an SBJ but no AR (top right), an AR but no SBJ (bottom left), and both an AR and an SBJ (bottom right) Linear regression lines are plotted in red, and the square of the correlation coefficient is provided in the upper right corner of each plot. Color shading denotes the maximum AR Scale observed at 37.5°N, 122.5°W, during the event.

References

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G.H., Curtis, J. & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Neiman, P. J., Hughes, M., Moore, B. J., Ralph, F. M., & Sukovich, E. M. (2013). Sierra Barrier Jets, Atmospheric Rivers, and Precipitation Characteristics in Northern California: A Composite Perspective Based on a Network of Wind Profilers. *Monthly Weather Review*, *141*(12), 4211–4233. https://doi.org/10.1175/MWR-D-13-00112.1

B.2.6 Landfalling ARs are often accompanied by large snow-level rises

Contributing Authors: Shawn Roj, Center for Western Weather and Water Extremes and Rachel Weihs, Center for Western Weather and Water Extremes

Data

Snow level climatology's from three vertically pointing snow level profilers were evaluated within the Yuba and Feather Rivers system: Downieville (DLA), New Bullards Bar (NBB), and Oroville (OVL). Field observations from the OVL Frequency-Modulated Continuous Wave (FMCW) snow level profiler (Johnston et al. 2017) that represent the calculated location of the brightband were downloaded from https://psl.noaa.gov/data/obs/datadisplay/. The S-Band snow level radar at OVL uses a brightband detection algorithm developed by White et al. 2002 and White et al. 2003. Measurements at OVL are averaged and saved at 10-minute intervals and span 1 January 2012 to 31 March 2023. Field observations from the DLA and NBB micro rain radar (MRR) snow level profilers that also represent the calculated location of the brightband were downloaded from the CW3E server. The brightband detection algorithm for the CW3E MRR profilers was modified from White et al. 2003 and Maahn and Kollias 2012. DLA data spans 1 December 2019 to 31 March 2023 and NBB data spans 16 December 2019 to 31 March 2023.

Measurement data from snow level radars are known to have some inaccuracies. Therefore, a quick quality control process was employed to minimize very large, short duration, snow level changes (SLC's) that may not be real or are not present long enough to affect the hydrology of a particular basin. Since the MRR sites are already averaged internally to hourly timesteps, the following process was only done for the 10-minute measurements at OVL. Step 1 was to calculate the centered 3-hour rolling average at each 10-minute timestep with a minimum of 6 observations needed. Step 2 was to calculate the absolute difference between each 10-minute observation and the 3-hour centered average. Step 3 was to omit the 10-minute measurement if it was more than 2 standard deviations from the 3-hour centered average. After this QC process was done, hourly median values of the snow level were calculated and used for the analysis herein.

Analysis

Detection of the brightband height from the snow level radars is only possible when there are melting hydrometeors present above each site. Therefore, the data are representative of precipitation events. Drought years can influence the distribution of snow levels if, for instance, there was a small number of all warm or all cold storms that affected a particular location that year. Therefore, caution should be used when the distribution suggests a mean snow level during years with less precipitation events.

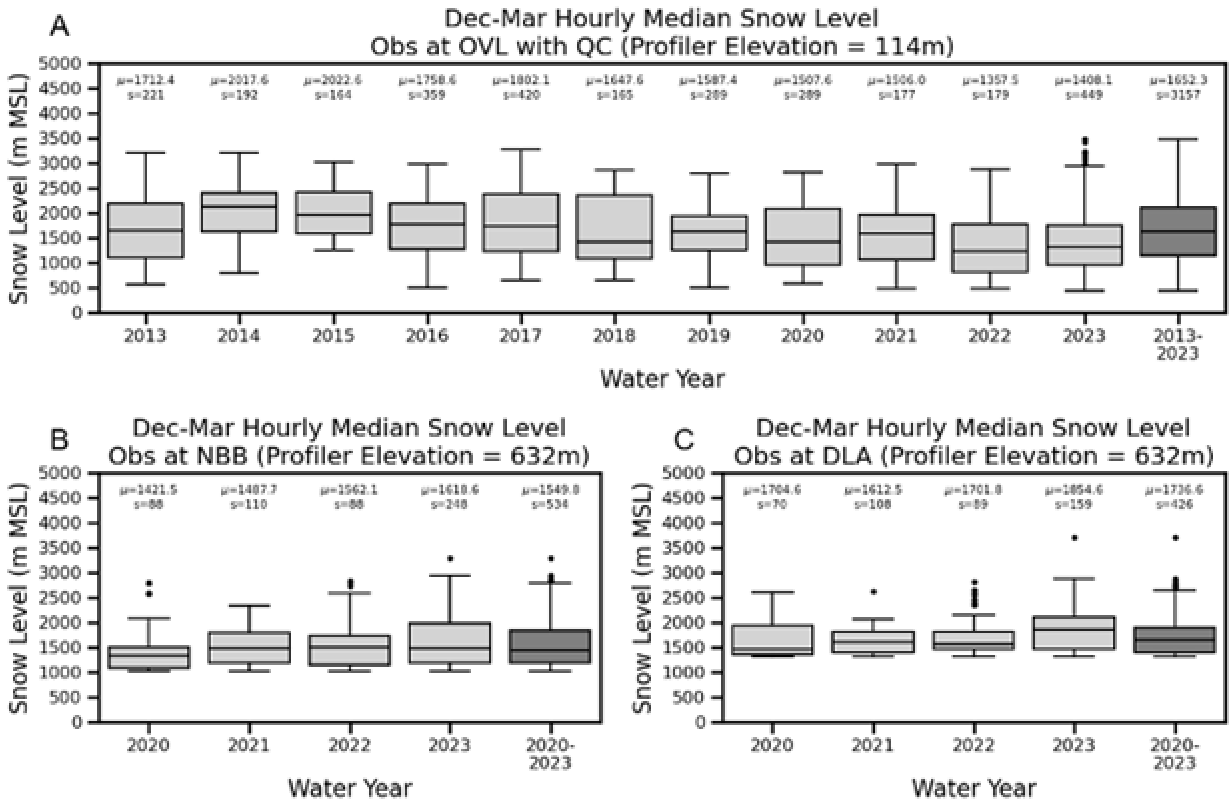


Figure B-11. Boxplots distributions of hourly median profiler snow levels by water year at OVL (1A), NBB (1B), and DLA (1C). The darker shaded boxplot represents the distribution for each site's entire period of record.

Figure B-11 shows boxplot distributions of hourly snow levels by water year and for the sites full period of record at each of the three profiler sites. Mean snow levels for the period of record at OVL, NBB, and DLA are 1652 m, 1549.8 m, and 1736.6 m, respectively. Mean snow levels at OVL show large variability between water years with a maximum mean snow level of 2022.6 m in water year 2015 and a minimum mean snow level of 1408.1 m in water year 2023. Not only was water year 2015 the 4th year of a major drought in California, but the storms that did impact the state were warm and exhibited higher snow levels that exacerbated snow drought conditions.

Water year 2023 was a record-breaking year for much of the U.S. West for precipitation and was also cooler than normal. This combination led to copious amounts of snow at upper elevations and much lower elevations where snow accumulations are less common. The distribution of snow levels at NBB and DLA exhibit less variation over the period of record at each site; 1421.5 m – 1618.6 m and 1612.5 m – 1854.6 m, respectively. In addition, both NBB and DLA exhibit more outliers likely due to their smaller sample sizes.

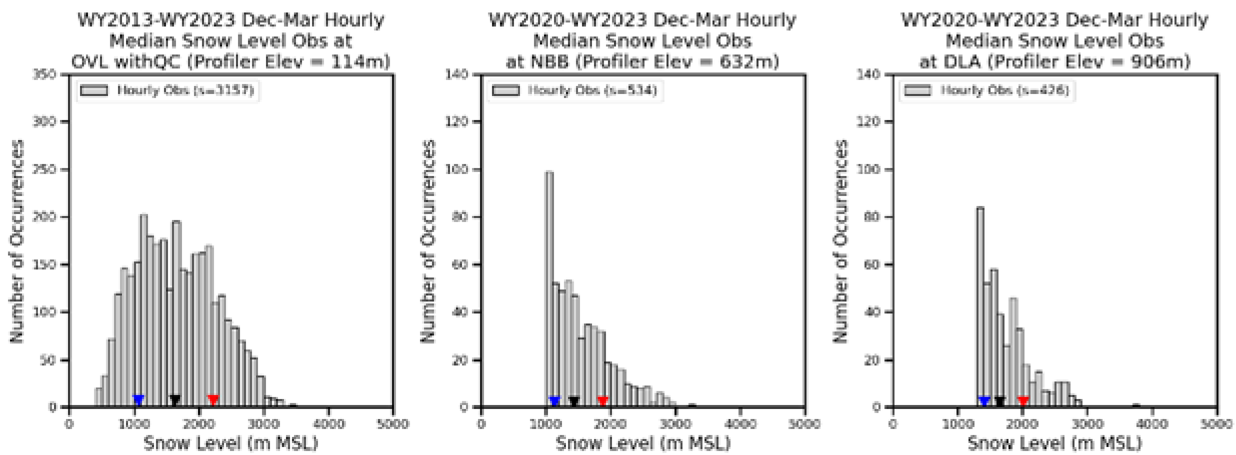


Figure B-12. Profiler snow level distribution histograms at OVL, NBB, and DLA, from left to right. The blue, black, and red triangles represent the 20th percentile, median, and 80th percentile, respectively, of hour median snow levels. Elevation bins are 100 m.

Figure B-12 has the same data as figure 1 but is now shown as histogram distributions at each site with the 20th percentile, median value, and 80th percentile of hourly median snow levels shown for the entire period of record at each site. The NBB and DLA histograms exhibit a sharp end at lower elevations. This is partly due to the profiler being located at an elevation high enough to receive snow, and therefore absent of a detectable brightband, and the resolution of the MRR close to the surface.

The MRRs are limited to discrete 'levels' of measurement. Due to this vertical resolution limitation, data from the two levels closest to the surface are omitted. This means that data within approximately 200 m of the surface are unavailable, thereby limiting the number of possible data points. The FMCW has a vertical resolution of approximately 40 m, and I am unaware of data that is omitted close to the surface. In addition, OVL is at a much lower elevation and does not often experience snow levels below the profiler elevation.

The distribution of hourly snow level rises (SLRs) and hourly snow level falls (SLFs) were also analyzed at each site. Figure B-13 shows histograms of hourly SLC's, where positive (negative) values represent hourly rises (falls). The distribution is even at OVL where 80% of the data on either side of zero lie between -110 m and 117.5 m with a median value of -2 m. At both NBB and DLA, these values are -150 m to 200 m with a median value of -50 m. This can be misleading though and there is probably room for an improved algorithm, but this mostly has to do with the resolution of the MRRs and the discrete elevation levels explained above.

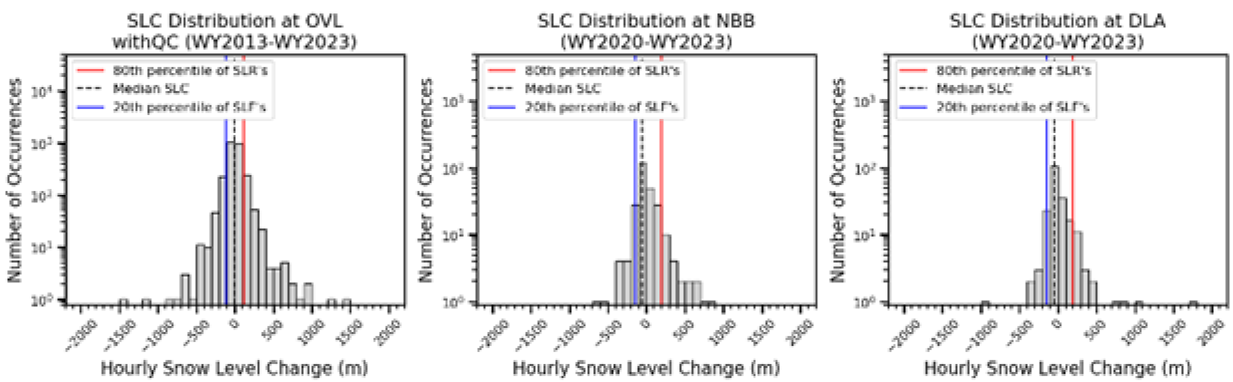


Figure B-13. Histograms of SLC's at OVL, NBB, and DLA, from left to right. The blue, dashed black, and red lines represent the 80^{th} percentile of hourly SLRs, the median hourly SLC's, and the 20^{th} percentile of hourly SLFs, respectively. Elevation bins are 100 m.

Because atmospheric rivers (ARs) are typically associated with the warm sector of an extratropical cyclone (Ralph et al., 2017) snow levels tend to be higher than non-AR related storms (Kim et al. 2013). The previous figures have shown the distribution of snow levels during precipitation events. For this analysis we wanted to better understand SLC's during landfalling ARs. We plotted scatter plots of hourly integrated water vapor transport (IVT) magnitude vs. hourly integrated water vapor (IWV) with dots shaded as hourly SLC's (Figure 4).

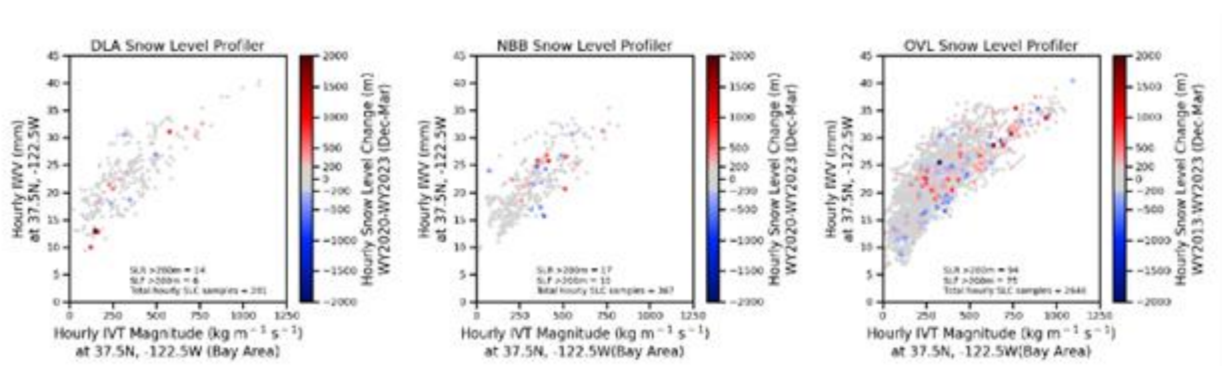


Figure B-14. Scatterplots of hourly IVT magnitude at a point near the Bay Area vs. hourly IWV at a point near the Bay Area, colored by hourly SLCs at DLA NBB, and OVL, from left to right

For clarity, only hourly SLCs greater than 200 m are shaded as red or blue, otherwise they are gray. There are very few SLCs greater than 200 m at both DLA and NBB which makes it difficult to identify any real patterns. However, the greatest SLCs tend to be SLRs at all sites. There appears to be some clustering of SLR's at OVL. When compared to the gray scatter points, SLR’s seem to occur during higher hourly IVT magnitudes and occur more often when IVT magnitudes are near 250 units and IWV values are near 20 mm.

Because of a lack of any real patterns and since the number of larger SLRs almost equals the number of SLFs, I then added a temporal dimension to the scatterplot. Figure B-15 has the same variables plotted but is now looking at 3-hourly changes of each. There is a positive correlation at OVL between increasing (decreasing) IVT magnitude, increasing (decreasing) IWV and snow levels rises (falls) (Figure B-15A and B-15B). Figures B-15A and B-15B are identical but B-15B has the SLFs plotted above the SLR’s for clarity. Again, only hourly SLCs greater than 200 m are shaded as red or blue, otherwise they are gray. This pattern is much less apparent at NBB and DLA due to much smaller sample sizes at each site (Figure B-15C and B-15D). At DLA, larger SLF’s appear to be like OVL, but larger SLR’s tend to occur more often when IVT magnitude and IWV decrease. At NBB, larger SLF’s also appear to be like OVL, but larger SLR’s seem to occur equally as often when IVT magnitude and IWV rise or fall.

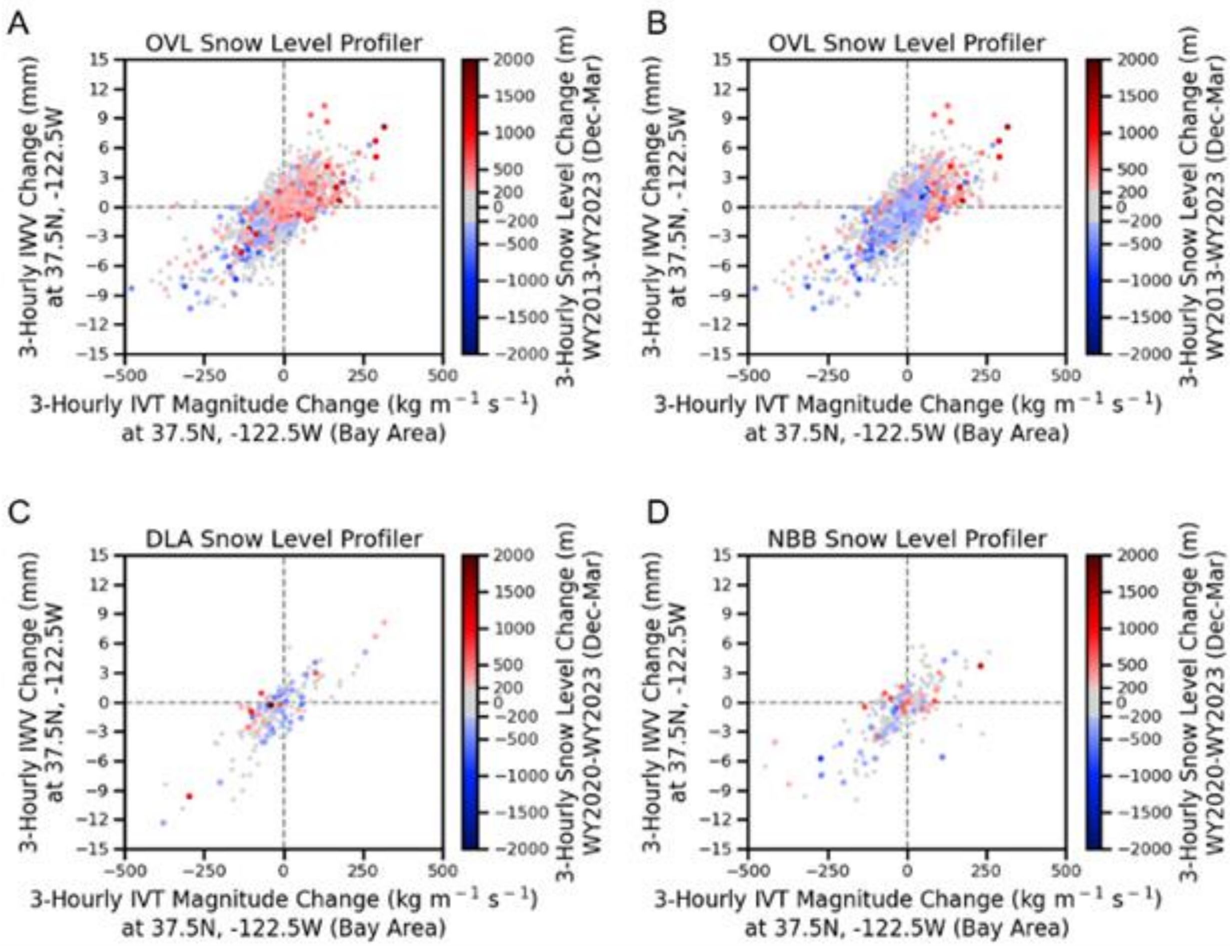


Figure B-15. Scatterplots of 3-hourly IVT magnitude change at a point near the Bay Area vs. 3-hourly IWV change at a point near the Bay Area, colored by 3-hourly SLCs at A) and B), OVL, C) DLA, and D) NBB. Figures B-15A and B-15B are identical but 4B has the decreasing 3-hourly snow level samples plotted above the increasing 3-hourly snow level samples for clarity. The color of each dot represents the corresponding 3-hourly SLC where SLFs >200 m and SLRs < 200 m are grey.

Finally, we looked at the relationship between hourly IVT magnitude, hourly IWV, and non-zero hourly precipitation from the meteorology stations co-located with each profiler (Figure B-16). This time, for clarity, samples with hourly precipitation < 5 mm are grey dots. At all three sites, the largest hourly precipitation events are not necessarily associated with the largest IVT magnitude and IWV. In comparison to NBB and OVL, DLA tends to receive more hourly precipitation when IVT magnitudes are relatively larger than IWV.

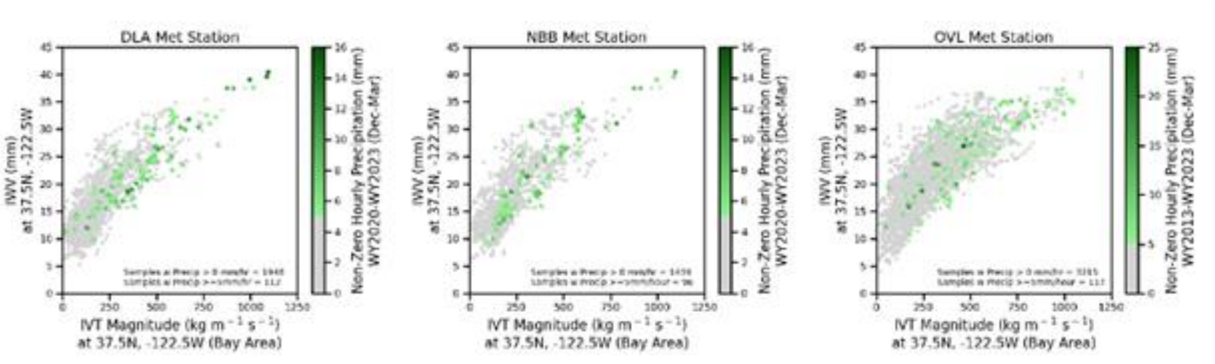


Figure B-16. Scatterplots of hourly IVT magnitude at a point near the Bay Area vs. hourly IWV at a point near the Bay Area, colored by hourly non-zero precipitation at DLA NBB, and OVL, from left to right

References

Johnston, P. E., Jordan, J. R., White, A. B., Carter, D. A., Costa, D. M., & Ayers, T. E. (2017). The NOAA FM-CW snow-level radar. Journal of Atmospheric and Oceanic Technology, 34(2), 249-267.

Kim, J., D. E. Waliser, P. J. Neiman, B. Guan, J.-M. Ryoo, and G. A. Wick (2013), Effects of atmospheric river landfalls on the cold season precipitation in California, Clim. Dyn., 40, 465–474, doi:10.1007/s00382-012-1322-3.

Maahn, M., & Kollias, P. (2012). Improved Micro Rain Radar snow measurements using Doppler spectra post-processing. Atmospheric Measurement Techniques, 5(11), 2661-2673.

Ralph, F. M., Iacobellis, S. F., Neiman, P. J., Cordeira, J. M., Spackman, J. R., Waliser, D. E., ... & Fairall, C. (2017). Dropsonde observations of total integrated water vapor transport within North Pacific atmospheric rivers. Journal of Hydrometeorology, 18(9), 2577-2596.

White, A. B., Gottas, D. J., Strem, E. T., Ralph, F. M., & Neiman, P. J. (2002). An automated brightband height detection algorithm for use with Doppler radar spectral moments. Journal of Atmospheric and Oceanic Technology, 19(5), 687-697.

White, A. B., Gottas, D. J., Ralph, F. M., & Neiman, P. J. (2003). U.S. Patent No. 6,615,140. Washington, DC: U.S. Patent and Trademark Office.

B.2.7 Maximum hourly precipitation rates are higher with landfalling ARs as compared to non-ARs

Contributing Authors: Samuel Bartlett, Center for Western Weather and Water Extremes

Data

A climatology of hourly precipitation observations from the (CDEC) database was gathered for a 21-year period between 00 UTC 01 Jan 2002 - 23 UTC 31 Jan 2023 at 28 stations within the Yuba & Feather watersheds. These stations consisted of a 9 in the North Fork Feather (NFF), 5 in the East Branch North Fork Feather (EBNFF), 6 in the Middle Fork Feather (MFF), and 8 in the Upper Yuba (UYB).

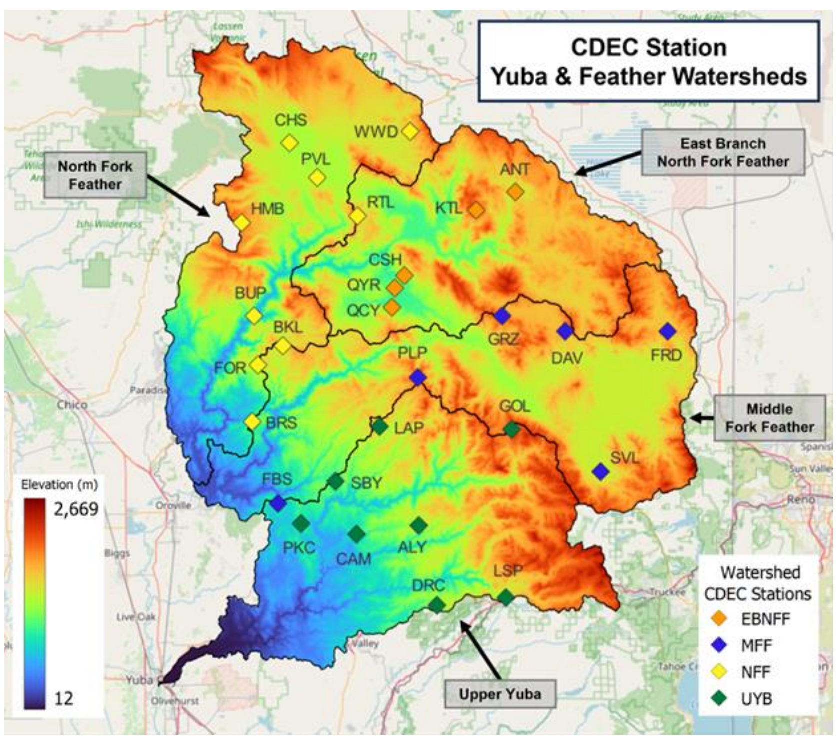


Figure B-17. *Yuba & Feather HUC-8 watershed boundaries (black lines), CDEC observation sites with precipitation observations during a majority of the study period (points, colored by sub watershed), and 30-meter resolution topographic data from the NASA Shuttle Radar Topography Mission (SRTM accessed via QGIS Plugin) (shaded).*

Data Analysis

Raw observational data was downloaded via the CDEC online data exchange for stations within the Yuba and Feather watersheds that had substantial data coverage during the period of record for the present study. The original data is formatted in an hourly-reported water-year accumulation value for each station in the catalog. Once this data was acquired, a calculation was performed to turn this water-year accumulation value into hourly-accumulation values by analyzing the difference in accumulation between each hour.

QC Process

After the hourly accumulation values were calculated, a three-step quality control (QC) process was conducted to limit the amount of erroneous data in the dataset. **QC#1** removed any hourly accumulation value at a single station greater than 10 inches in one hour. A 10 in/hr rate was chosen as this is near the upper limit of what is possible in the atmosphere and above the NOAA Atlas 14 100-year return interval value for 60 minute observed precipitation in the region (Attached full map and inset). **QC#2** removed hourly observations at a single station if they were greater than 1 inch (25.4 mm) AND greater than the sum of all observations in the sub-watershed. This QC step was designed to eliminate single station erroneous values which are unlikely due to isolated, high single hourly precipitation observations. **QC#3** verifies that any hourly accumulation which was calculated for the original water year total is associated with a valid hourly observation, and not a missing value in the dataset. Across all stations in the dataset, the highest amount of missing and QC'ed out values occurred within the NFF sub-watershed at the RTL station, with 17,012 hours missing (9.2%) and 354 hours QC'ed out (0.19%), combined for a total of 9.4% of the data removed from the original dataset for this station.

Research Questions

1. What was the average event maximum hourly precipitation rate during No-AR vs AR events over each subwatershed?
2. What was the average event maximum hourly precipitation rate during AR1 & AR2 events vs AR3, AR4, AR5 events?
3. For all hourly precipitation observations greater than zero during each event, what was the average hourly precipitation rate during No-AR vs AR events?
4. For all hourly precipitation observations greater than zero during each event, what was the average hourly precipitation rate during AR1 & AR2 events vs AR3, AR4, AR5 events?
5. What was the average percent of the total AR duration with observed precipitation values greater than zero during No-AR vs AR events?
6. What was the average percent of the total AR duration with observed precipitation values greater than zero during AR1 & AR2 events vs AR3, AR4, AR5 events?

7. Over the full POR, what is the 99th, 95th, and 90th percentile hourly precipitation value observed at each station and over each subwatershed?

Analysis #1

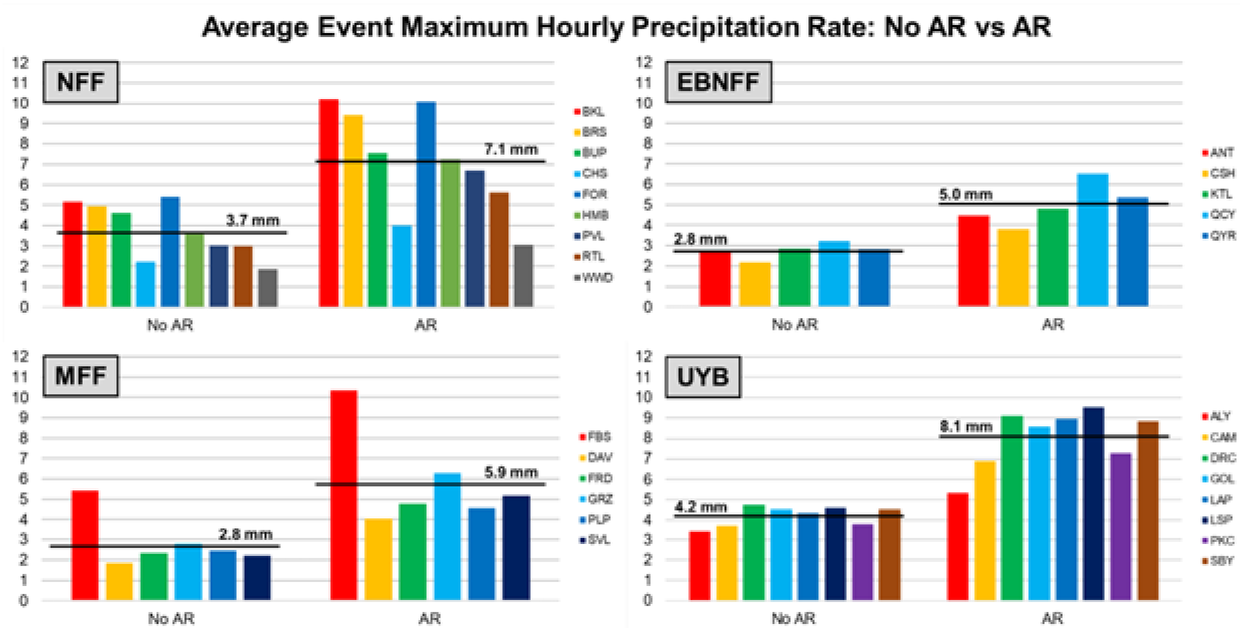


Figure B-18. Average maximum hourly precipitation rate observed during the top 90th percentile precipitation events in the Feathers and Yuba watersheds, plotted for the stations in each subwatershed (columns), with the average of all stations plotted (horizontal black lines), subset by No AR vs AR precipitation events.

Method: For all hours of each precipitation event, pull out the <u>maximum observed hourly precipitation rates</u>. Take all the maximum values for each event and average them for each station, this is shown as the <u>column value</u>. Take all these averages for each station, then average those to get a single value for each analysis in the subwatersheds which are represented by the <u>horizontal black lines</u>. This analysis was subset for precipitation events that had <u>No AR vs those that had and AR</u>

Analysis: On average, the maximum hourly precipitation rates during precipitation events is almost double during AR related precipitation vs No AR related precipitation across all subwatersheds. When comparing No AR events to AR events, in the NFF there is a 92% increase, EBNFF a 79% increase, MFF a 111% increase, and UYB a 93% increase, which averages out to 94% across the 4 sub watersheds.

NFF	No AR: 3.7 mm	AR: 7.1 mm	Percent Increase: 92%
EBNFF	No AR: 2.8 mm	AR: 5.0 mm	Percent Increase: 79%
MFF	No AR: 2.8 mm	AR: 5.9 mm	Percent Increase: 111%
UYB	No AR: 4.2 mm	AR: 8.1 mm	Percent Increase: 93%

Analysis #2

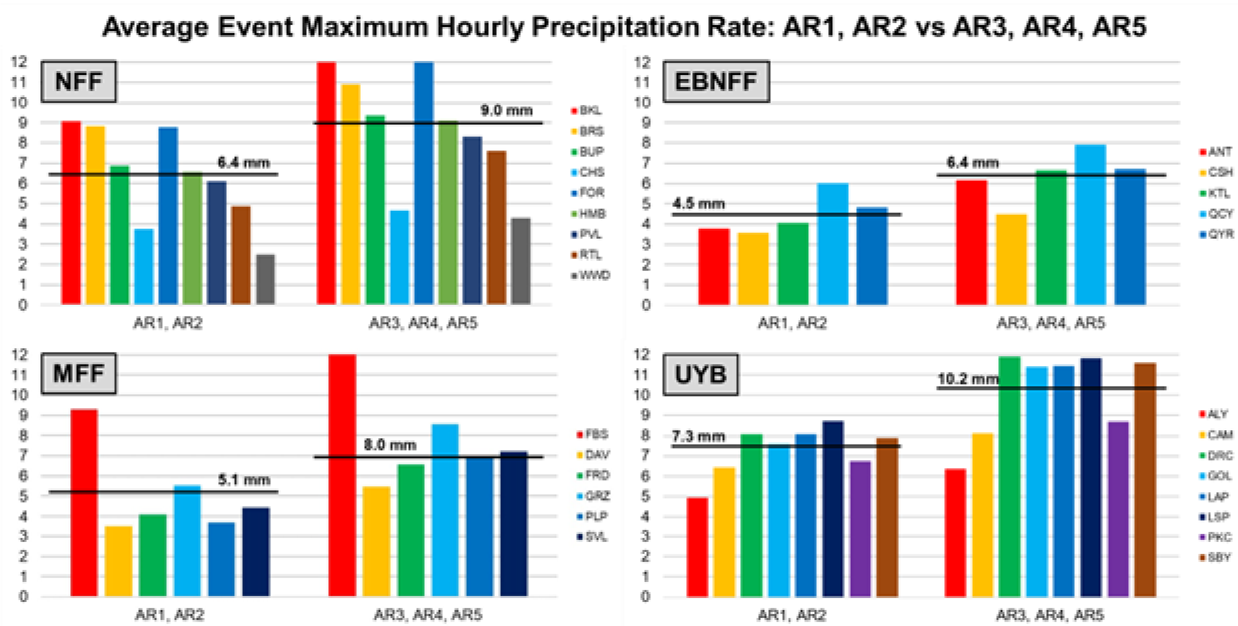


Figure B-19. Average maximum hourly precipitation rate observed during the top 90th percentile AR related precipitation events in the Feathers and Yuba watersheds, plotted for the stations in each subwatershed (columns), with the average of all stations plotted (horizontal black lines), subset by AR1, AR2 and AR3, AR4, AR5 ranked events.

Method: For all hours of each AR related precipitation event, pull out the maximum observed hourly precipitation rates. Take all the maximum values for each event and average them for each station, this is shown as the column value. Take all these averages for each station, then average those to get a single value for each analysis in the subwatersheds which are represented by the horizontal black lines. This analysis was subset for precipitation events that were ranked AR1, AR2 vs AR3, AR4, AR5 based on the Ralph et al. 2019 AR scale.

Analysis: As AR intensity increases from AR1, AR2 to AR3, AR4, AR5 the maximum hourly precipitation rates at each station increases by approximately 50% across stations in each sub-watershed. In the NFF there is a 41% increase, EBNFF a 42% increase, MFF a 56% increase, and UYB a 40% increase, which averages out to 45% across the 4 sub watersheds.

NFF	AR1, AR2: 6.4 mm	AR3, AR4, AR5: 9.0 mm	Percent Increase: 41%
EBNFF	AR1, AR2: 4.5 mm	AR3, AR4, AR5: 6.4 mm	Percent Increase: 42%
MFF	AR1, AR2: 5.1 mm	AR3, AR4, AR5: 8.0 mm	Percent Increase: 56%
UYB	AR1, AR2: 7.3 mm	AR3, AR4, AR5: 10.2 mm	Percent Increase: 40%

Analysis #3

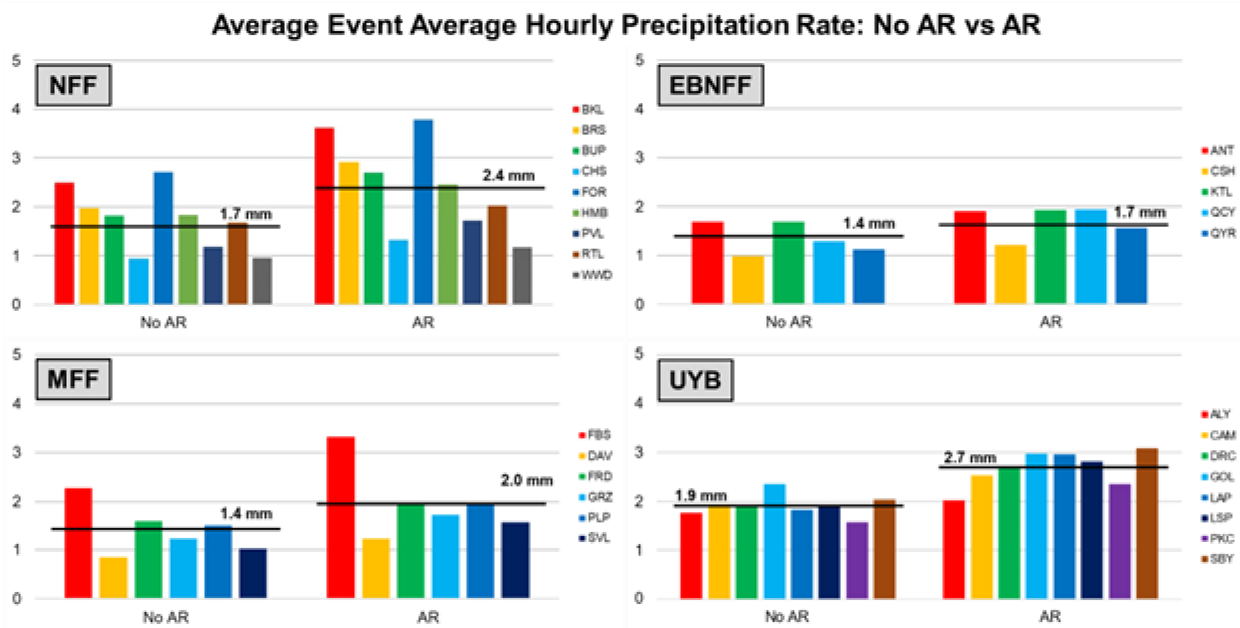


Figure B-20. Average hourly precipitation rate observed during the top 90th percentile precipitation events in the Feathers and Yuba watersheds, plotted for the stations in each subwatershed (columns), with the average of all stations plotted (horizontal black lines), subset by No AR vs AR precipitation events.

Method: Take the average of all hours of precipitation greater than zero during each event. Then average these event values across each station which is represented as the columns. Take all these averages for each station, then average those to get a single value for each analysis in the subwatersheds which are represented by the horizontal black lines. This analysis was subset for precipitation events that had No AR vs those that had and AR

Analysis: On average, the average hourly precipitation rates during precipitation events at stations in the Yuba and Feather watersheds is 37% higher during AR events as compared to No AR Events. The percent increase across stations in eac subwatershed are as follows, 41% increase in the NFF, 21% increase in the EBNFF, 43% increase in the MFF, and 42% increase in the UYB.

NFF	No AR: 1.7 mm	AR: 2.4 mm	Percent Increase: 41%
EBNFF	No AR: 1.4 mm	AR: 1.7 mm	Percent Increase: 21%
MFF	No AR: 1.4 mm	AR: 2.0 mm	Percent Increase: 43%
UYB	No AR: 1.9 mm	AR: 2.7 mm	Percent Increase: 42%

Analysis #4

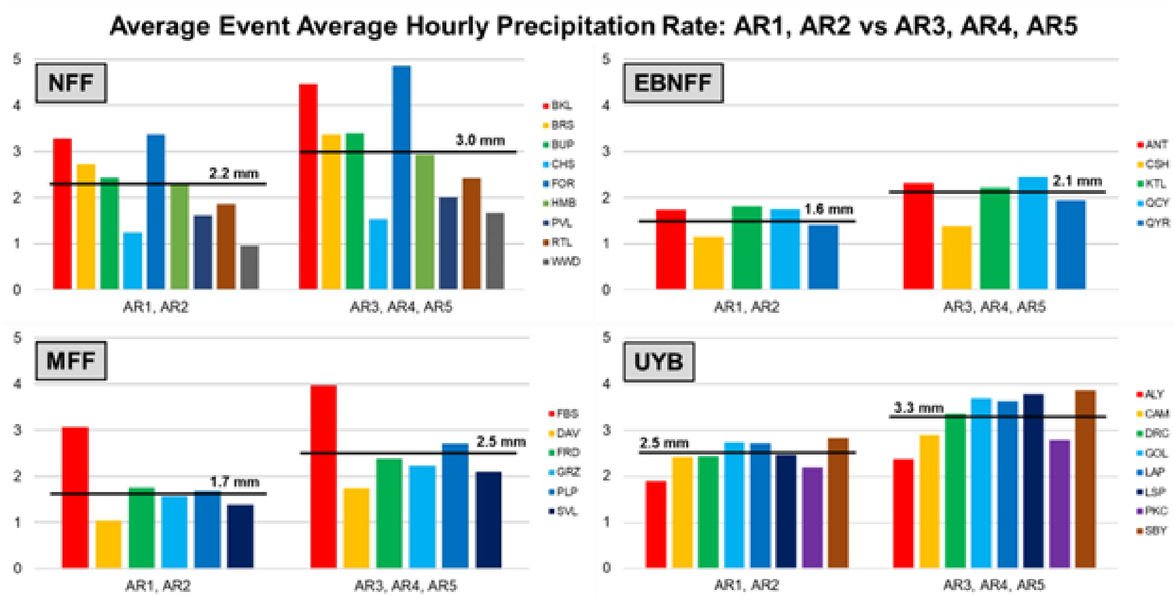


Figure B-21. Average hourly precipitation rate observed during the top 90th percentile AR related precipitation events in the Feathers and Yuba watersheds, plotted for the stations in each subwatershed (columns), with the average of all stations plotted (horizontal black lines), subset by AR1, AR2 and AR3, AR4, AR5 ranked events.

Method: Take the average of all hours of precipitation greater than zero during each AR related event. Then average these event values across each station which is represented as the columns. Take all these averages for each station, then average those to get a single value for each analysis in the subwatersheds which are represented by the horizontal black lines. This analysis was subset for precipitation events that were ranked AR1, AR2 vs AR3, AR4, AR5 based on the Ralph et al. 2019 AR scale.

Analysis: On average across the watershed, as AR intensity increases from AR1, AR2 to AR3, AR4, AR5 the event average hourly precipitation observations during strong ARs are 37% higher than weaker ARs. Across the subwatersheds, strong ARs have 36% higher average hourly precipitation rates, 31% higher in the EBNFF, 47% higher in the MFF, and 32% higher in the UYB.

NFF	AR1, AR2: 2.2 mm	AR3, AR4, AR5: 3.0 mm	Percent Increase: 36%
EBNFF	AR1, AR2: 1.6 mm	AR3, AR4, AR5: 2.1 mm	Percent Increase: 31%
MFF	AR1, AR2: 1.7 mm	AR3, AR4, AR5: 2.5 mm	Percent Increase: 47%
UYB	AR1, AR2: 2.5 mm	AR3, AR4, AR5: 3.3 mm	Percent Increase: 32%

Analysis #5

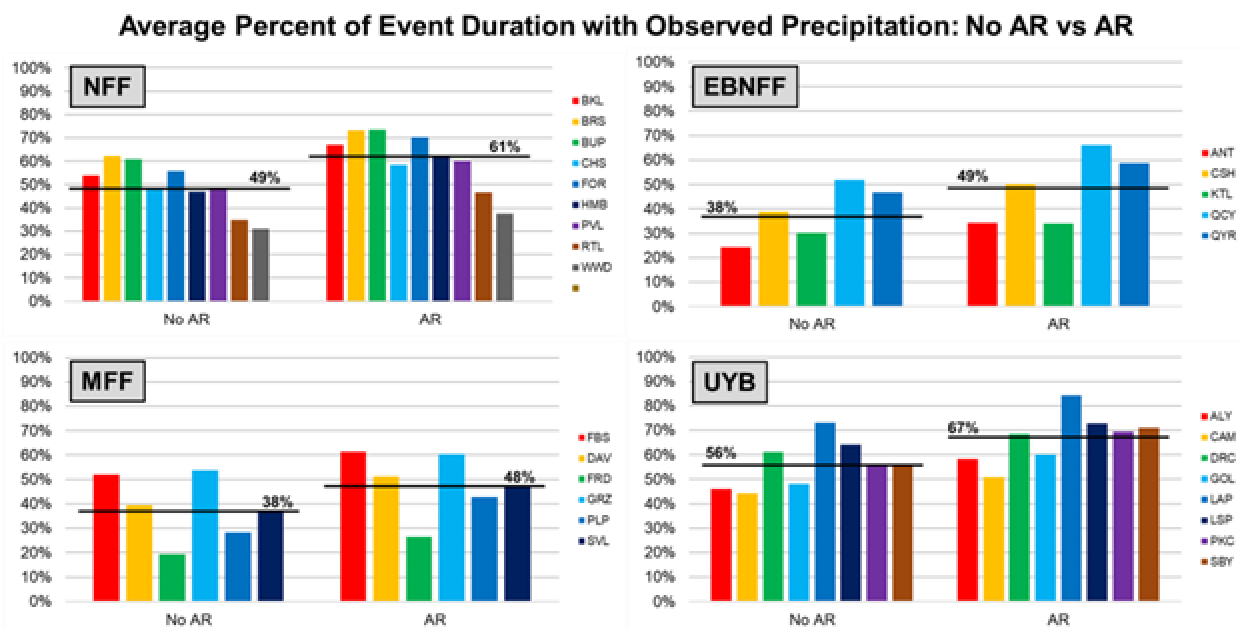


Figure B-22. Average percent of precipitation event duration with observed precipitation for each station in the Feather and Yuba watersheds (columns), with the average of all stations plotted (horizontal black lines), subset by No AR vs AR precipitation events.

Method: For each event, count the number of hours with precipitation observations greater than zero at each station. Divide that number by the event duration to establish a percent value. Average all these percentages for each station (columns), then average the station values across each subwatershed (black lines). This analysis was subset for precipitation events that had No AR vs those that had and AR

Analysis: On average across the 4 sub watersheds, the average percent of each event duration with observed precipitation increased by 25% when comparing No AR-to-AR events.

Broken down by subwatersheds, the percent increase is 24% in the NFF, 29% in the EBNFF, 26% in the MFF, and 20% in the UYB.

NFF	No AR: 49%	AR: 61%	Percent Increase: 24%
EBNFF	No AR: 38%	AR: 49%	Percent Increase: 29%
MFF	No AR: 38%	AR: 48%	Percent Increase: 26%
UYB	No AR: 56%	AR: 67%	Percent Increase: 20%

Analysis #6

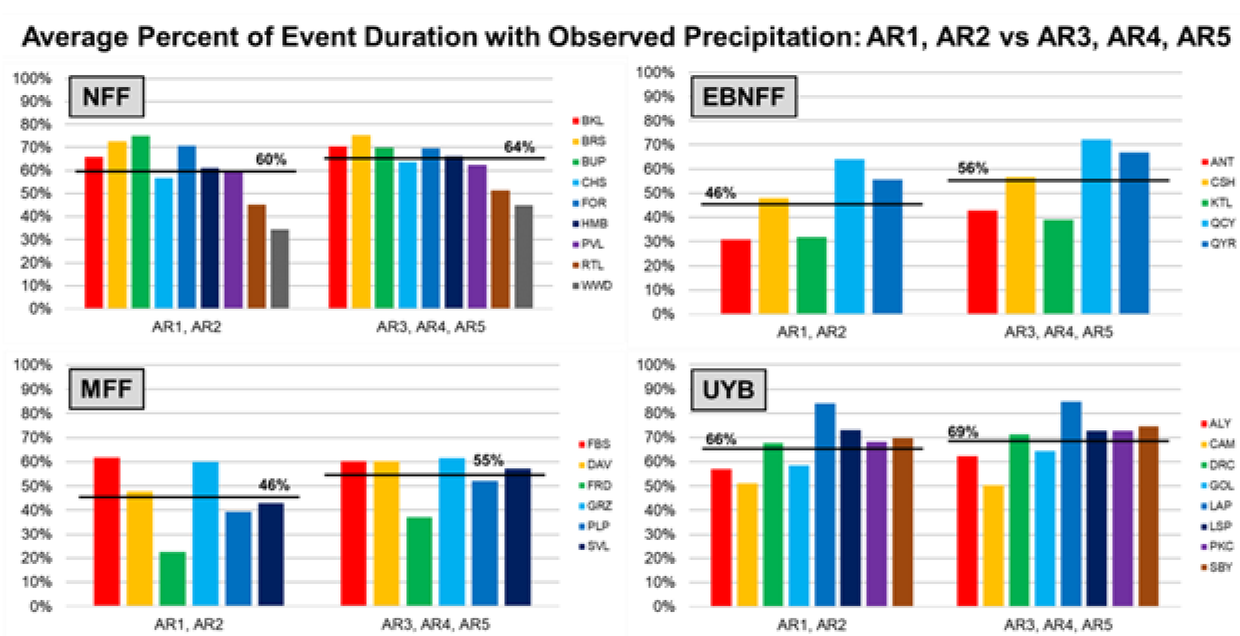


Figure B-23. Average percent of AR Related precipitation event duration with observed precipitation for each station in the Feather and Yuba watersheds (columns), with the average of all stations plotted (horizontal black lines), subset by AR1, AR2 and AR3, AR4, AR5 ranked events.

Method: For each AR Related event, count the number of hours with precipitation observations greater than zero at each station. Divide that number by the event duration to establish a percent value. Average all these percentages for each station (columns), then average the station values across each subwatershed (black lines). This analysis was subset for precipitation events that were ranked AR1, AR2 vs AR3, AR4, AR5 based on the Ralph et al. 2019 AR scale.

Analysis: As AR intensity increases from AR1, AR2 to AR3, AR4, AR5 the average duration of precipitation increases by 13% across each subwatershed, with a percent increase of 7% in the NFF, 22% increase in the EBNFF, 20% in the MFF, and 5% in the UYB. Stronger ARs have marginally longer precipitation durations as compared to weak ARs, but this analysis does not consider the difference in duration of weak events as compared to longer, strong AR events. Some form of precipitation duration percentage normalized for event duration would be required to draw additional conclusions.

NFF AR1, AR2: 60% AR3, AR4, AR5: 64% Percent Increase: 7%

EBNFF AR1, AR2: 46% AR3, AR4, AR5: 56% Percent Increase: 22%

MFF AR1, AR2: 46% AR3, AR4, AR5: 55% Percent Increase: 20%

UYB AR1, AR2: 66% AR3, AR4, AR5: 69% Percent Increase: 5%

Analysis #7

Station	Watershed	Elevation (ft)	99th	95th	90th
BKL	NFF	5,873	13.21	7.85	6.10
BRS	NFF	3,560	12.19	7.58	5.59
BUP	NFF	1,760	9.65	6.10	4.83
CHS	NFF	4,525	5.84	3.30	2.54
FOR	NFF	5,202	13.21	8.13	6.10
HMB	NFF	6,500	8.13	5.08	4.06
PVL	NFF	4,520	8.13	4.32	3.05
RTL	NFF	6,210	7.11	4.06	3.05
WWD	NFF	5,150	5.84	3.56	2.54
ANT	EBNFF	4,960	9.14	4.06	3.05
CSH	EBNFF	4,520	5.84	3.56	2.79
KTL	EBNFF	7,300	7.11	4.06	3.05
QCY	EBNFF	3,408	8.64	5.33	3.81
QYR	EBNFF	3,500	7.59	4.32	3.30
FBS	MFF	2,840	14.22	8.13	6.10
DAV	MFF	5,768	6.60	3.30	2.29
FRD	MFF	5,517	10.16	5.08	3.05
GRZ	MFF	6,900	7.11	3.81	2.54
PLP	MFF	6,800	7.11	4.06	3.05
SVL	MFF	4,975	8.13	4.57	3.05
ALY	UYB	4,957	9.14	5.08	4.06
CAM	UYB	2,755	10.16	6.10	5.08
DRC	UYB	4,455	11.94	6.86	5.08
GOL	UYB	6,750	12.19	7.11	5.08
LAP	UYB	4,980	10.41	6.35	4.57
LSP	UYB	5,156	10.92	6.35	4.83
PKC	UYB	3,714	9.65	6.10	4.57
SBY	UYB	3,810	11.43	7.11	5.33

Figure B-24. The 99th, 95th, and 90th percentile threshold 1-hour precipitation observations for stations in the NFF, :EBNFF, MFF, and UYB watersheds. Listed by three letter station identifiers within each subwatershed, elevation (ft), and statistical threshold values (mm).

Method: For all hourly precipitation observations greater than zero during the full period of record, compute the 99th, 95th, and 90th percentile threshold values for each station within the subwatershed.

Analysis: An analysis of the relationship between precipitation observations and elevation yields no significant results in terms of precipitation event maximum hourly observation, event average hourly observation, or event duration with precipitation observed.

Analysis #7a

Station	Watershed	Elevation (ft)	99th	95th	90th
BUP	NFF	1,760	9.65	6.10	4.83
CAM	UYB	2,755	10.16	6.10	5.08
FBS	MFF	2,840	14.22	8.13	6.10
QCY	EBNFF	3,408	8.64	5.33	3.81
QYR	EBNFF	3,500	7.59	4.32	3.30
BRS	NFF	3,560	12.19	7.58	5.59
PKC	UYB	3,714	9.65	6.10	4.57
SBY	UYB	3,810	11.43	7.11	5.33
DRC	UYB	4,455	11.94	6.86	5.08
PVL	NFF	4,520	8.13	4.32	3.05
CSH	EBNFF	4,520	5.84	3.56	2.79
CHS	NFF	4,525	5.84	3.30	2.54
ALY	UYB	4,957	9.14	5.08	4.06
ANT	EBNFF	4,960	9.14	4.06	3.05
SVL	MFF	4,975	8.13	4.57	3.05
LAP	UYB	4,980	10.41	6.35	4.57
WWD	NFF	5,150	5.84	3.56	2.54
LSP	UYB	5,156	10.92	6.35	4.83
FOR	NFF	5,202	13.21	8.13	6.10
FRD	MFF	5,517	10.16	5.08	3.05
DAV	MFF	5,768	6.60	3.30	2.29
BKL	NFF	5,873	13.21	7.85	6.10
RTL	NFF	6,210	7.11	4.06	3.05
HMB	NFF	6,500	8.13	5.08	4.06
GOL	UYB	6,750	12.19	7.11	5.08
PLP	MFF	6,800	7.11	4.06	3.05
GRZ	MFF	6,900	7.11	3.81	2.54
KTL	EBNFF	7,300	7.11	4.06	3.05

Figure B-24. As in Figure B-23, except sorted from lowest elevation to highest elevation, shaded by subwatershed.

NFF	BKL_NFF_5873	BRS_NFF_3560	BUP_NFF_1760	CHS_NFF_4525	FOR_NFF_5202	HMB_NFF_650(	PVL_NFF_4520	RTL_NFF_6210	WWD_NFF_5150
Missing_hrs	8281	3909	6988	1945	8043	12962	2616	17012	4449
Percent Missing	4.48%	2.11%	3.78%	1.05%	4.35%	7.01%	1.42%	9.20%	2.41%
QCed_hrs	124	25	103	18	73	54	55	354	8
Percent Qced	0.07%	0.01%	0.06%	0.01%	0.04%	0.03%	0.03%	0.19%	0.00%
Total_-9999.99_hrs	8405	3934	7091	1963	8116	13016	2671	17366	4457
Per_MISS_QC_hrs	4.55%	2.13%	3.84%	1.06%	4.39%	7.04%	1.45%	9.40%	2.41%
EBNFF	ANT_EBNFF_49	CSH_EBNFF_452	KTL_EBNFF_730	QCY_EBNFF_34(	QYR_EBNFF_3500				
Missing_hrs	1845	12186	7897	2170	9633				
Percent Missing	1.00%	6.59%	4.27%	1.17%	5.21%				
QCed_hrs	150	16	33	13	13				
Percent Qced	0.08%	0.01%	0.02%	0.01%	0.01%				Period of Record
Total_-9999.99_hrs	1995	12202	7930	2183	9646				N=184,824 hrs
Per_MISS_QC_hrs	1.08%	6.60%	4.29%	1.18%	5.22%				
MFF	FBS_MFF_2840	DAV_MFF_5768	FRD_MFF_5517	GRZ_MFF_6900	PLP_MFF_6800	SVL_MFF_4975			
Missing_hrs	11194	7286	5957	15024	7576	2967			
Percent Missing	6.06%	3.94%	3.22%	8.13%	4.10%	1.61%			
QCed_hrs	107	14	118	67	26	190			
Percent Qced	0.06%	0.01%	0.06%	0.04%	0.01%	0.10%			
Total_-9999.99_hrs	11301	7300	6087	15091	7602	3157			
Per_MISS_QC_hrs	6.11%	3.95%	3.29%	8.17%	4.11%	1.71%			
UYB	ALY_UYB_4957	CAM_UYB_2755	DRC_UYB_4455	GOL_UYB_6750	LAP_UYB_4980	LSP_UYB_5156	PKC_UYB_3714	SBY_UYB_3810	
Missing_hrs	6951	3650	7528	3963	2662	3104	16882	5567	
Percent Missing	3.76%	1.97%	4.07%	2.14%	1.44%	1.68%	9.13%	3.01%	
QCed_hrs	42	141	127	49	37	154	31	48	
Percent Qced	0.02%	0.08%	0.07%	0.03%	0.02%	0.08%	0.02%	0.03%	
Total_-9999.99_hrs	6993	3791	7655	4012	2699	3258	16913	5615	
Per_MISS_QC_hrs	3.78%	2.05%	4.14%	2.17%	1.46%	1.76%	9.15%	3.04%	

Figure B-25. *Number of missing, QC'ed out, and sum of missing & QC'ed out hourly observation at each station, with the percentage of the full period of record listed below each value. Percentage values are calculated by dividing the counts for each category by the full study period duration of 184,824 hours.*

# of Events	Feathers	Yuba
No AR	584	556
AR	202	198
AR1	91	94
AR2	56	53
AR3	42	38
AR4	12	12
AR5	1	1
AR1, AR2	147	147
AR3, AR4, AR5	55	51

Figure B-26. *Number of events in each category in the present study for No AR and AR events (grey), AR1, AR2, AR3, AR4, and AR5 events (rainbow), and AR1 & AR2, and AR3, AR4, AR5 events (pinks) from the AR catalogs developed for the Feather and Yuba watersheds.*

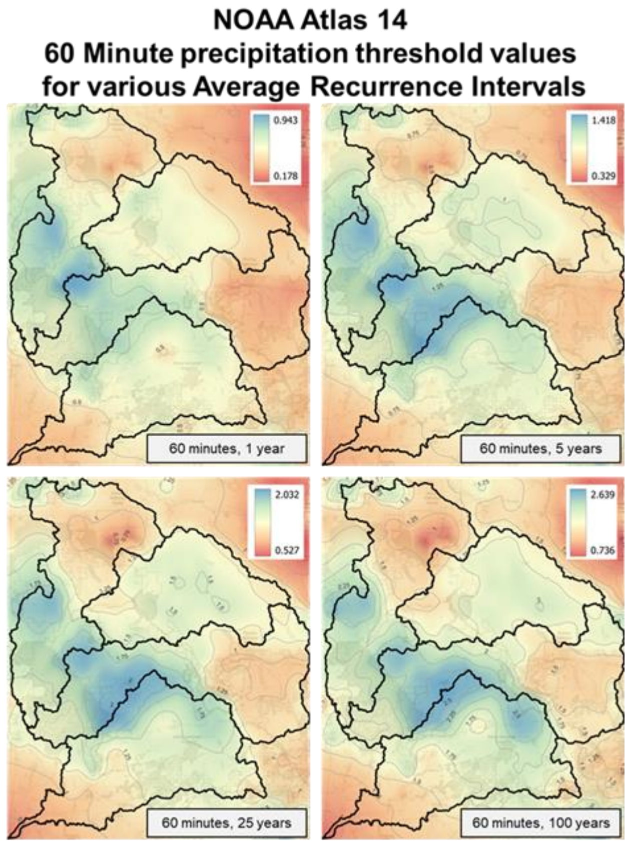


Figure B-27. *NOAA Atlas 14 60-minute precipitation threshold values (shaded, contoured) for 1-year (top left), 5-year (top right), 25-year (bottom left), and 100-year (bottom right) recurrence intervals over the Yuba and Feather watersheds (black lines).*

B.2.8 Extreme hourly precipitation rates are not necessarily driven by NCFRs

Contributing Authors: Jon Rutz, Center for Western Weather and Water Extremes and Matthew Steen, Center for Western Weather and Water Extremes

We used hourly precipitation data and archived radar imagery to quantify the fraction of extreme precipitation events associated with narrow cold-frontal rain bands (NCFRs) across the Yuba-Feather (Y-F) Watersheds. First, quality control measures were performed on the hourly precipitation data obtained from the California Data Exchange Center (CDEC) to ensure only high-quality data for the analysis. Next, we identify the extreme (i.e., top 25) precipitation events from each sub-watershed and merge these lists, resulting in 55 unique events due to overlapping events. We can manually examine NEXRAD radar data from Sacramento, CA (DAX) to assess whether an NCFR impacted the Y-F Watersheds, which fall roughly within ~39.3 N to ~40.3 N and ~ -120.2 W to ~ -121.4 W, using the following rubric (Example provided in Figure B-28):

1. **Yes:** extremely high confidence that an NCFR exists and overlaps some portion of the watersheds
2. **No:** extremely high confidence that there is no NCFR present or that if present, the NCFR does not overlap some portion of the watersheds during this hour
3. **Maybe:** either 1) a feature with low-confidence identification as an NCFR that overlaps some portion of the watersheds, or 2) a feature with high-confidence identification as an NCFR that very nearly overlaps some portion of the watersheds, or 3) some combination of the uncertainties above

After accounting for missing radar data, 49 unique events were analyzed (approximately 2.5 events per year, but these events are not equally distributed across years). For these events, 9 (16%) are defined as yes, 39 (71%) as no, and 1 (2%) as maybe. In other words, the chance of any extreme precipitation event (i.e.) being associated with an NCFR is ~16%. We caution that while this appears to assign a relatively low overlap between NCFRs and extreme hourly precipitation in the Y-F Watersheds, the real headline is that extreme precipitation here does not require an NCFR – orographic enhancement and other processes play key roles in most events. However, what this study does not answer is how often extreme hourly precipitation occurs when an NCFR is present. Answering that question is beyond the scope of this study and requires radar analysis of all precipitation events to determine how many NCFRs are not associated with extreme hourly precipitation.

For NCFRs: (aka how we got the dates)

Precipitation QC:

1. Acquire CDEC data sheets for each of the subwatershed from Sam Bartlett. This data had been previously QC'd by Sam, we were looking to take it a step further. The four sub-watersheds are East Branch North Fork Feather (EBNFF), Middle Fork Feather (MFF), North Fork Feather (NFF) and Upper Yuba (UYB).
2. From each data sheet, I looked at each set of hourly precipitation observations.
3. If any single station observation in that hour was assigned the '-9999.99' missing tag, it will now be assigned a zero. This will assist with a QC step later.
4. If half or more of the station hourly observations are 0, this hour of precipitation will not be considered and is 'thrown out'. This equates to the following number of stations:

   EBNFF >= 3 stations

   MFF >= 3 stations

NFF >= 5 stations

UYB >= 4 stations

We decided on this step to help eliminate hours of precipitation where most of the watershed was not impacted in some way. This also effectively removed all hours in which there was no measured precipitation, which helped limit the number of hours of data in consideration.

5. Next, we sum up the station observations for the hour and calculate the percentage contribution of each station to the sum.
6. If any one station's contribution is 50% or greater, this hour of precipitation is not considered and is also 'thrown out'. This step was taken to remove hours in which the precipitation at one station seems unrealistically high relative to other stations (note that this step may remove some real events but is worthwhile to increase our confidence in the remaining events). Hours of precipitation that have passed both checks are then considered for analysis

Event Selection:

1. Events needed to be selected and defined prior to further analysis.
2. For each subwatershed, the remaining hours of precipitation were sorted by their total precipitation. The goal was to analyze the top 25 events per watershed.
3. For this case, an event is defined as any unique hour(s) of precipitation. For an hour of precipitation to be grouped together and be defined as an event there must be a ≤ 6-hour gap between observations.
    1. ex/ Precipitation on 09/07 at 13Z and 18Z would be grouped together, but precipitation at 13Z and 21Z would not be.
4. Once the top 25 were identified for each subwatershed, a master list of events was created by combining the four lists. This resulted in 55 unique events across the four subwatershed due to overlapping events

Radar Analysis:

1. For each masterlist event, short-range base reflectivity at 0.5° tilt angle and short-range composite reflectivity were gathered for the hours of observation as well as the 6 hours before and after from the NCEI NEXRAD Radar Archive (NEXRAD Data Inventory Search | National Centers for Environmental Information (noaa.gov)). The Sacramento, CA radar, KDAX, was selected as the radar of choice due to its central location in Northern California, upstream of the Yuba-Feather Watersheds.
    1. These radar products were chosen based on Orla-Barile et al. 2021 methodology for their climatology of NCFRs in Southern California (A Climatology of Narrow Cold-Frontal Rainbands in Southern California - Orla-Barile - 2022 - Geophysical Research Letters - Wiley Online Library)

    2. If either product was unavailable, the long-range alternative was selected instead. If neither were available then that event would not be analyzed.
2. For each event, the goal was to identify if there was an NCFR impacting the watersheds. The subwatersheds fall between a box of latitude ~39.3° N to ~40.3° N and longitude ~ -120.2 W to ~ -121.4 W.
    1. During the top hours identified for the event, if any portion of the NCFR passed through this box at all during a scan in the hour(s), it would count for the purpose of this analysis.
    2. A feature would be identified as an NCFR if a region of ≥ 40 dbz was found to have a length to width ratio of 5:1 or greater.
    3. During the hour(s) of the event, a yes/no/maybe would be assigned to the hour based on the presence of an NCFR of the definition above.
        1. Yes is a definite NCFR with any portion within the boundaries of the watershed
        2. Maybe is a feature that might be an NCFR that passes within the boundaries of the watershed or if only a small portion of the NCFR passes through the boundaries
        3. No is there is no NCFR present or a present NCFR does not pass near the watershed boundaries
3. A similar methodology is applied to the six hours before and after. If there is a maybe or yes during the period, the specific time is noted not just blankly noted as maybe or yes

For Precipitation Events:

1. Beginning with the QC'd precipitation dates used in the NCFR analysis, we wanted to locate the top percentile hourly precipitation in each watershed.
2. How we elected to look at this was not directly based on total precipitation but to look at the average percentile hourly precipitation across the stations in the watershed. We hoped that by averaging precipitation percentiles, stations that consistently received greater precipitation totals did not exhibit more influence on the events that would be selected.
    a. For example, the second largest hourly precipitation total across the NFF subbasin occurred on 2/4/17 at 14Z where 122 mm was measured by the station. 98.552 mm came from two of the nine stations (54.864 at BRS and 43.688 at FOR). While this is the second highest total, when averaging the percentiles across the subwatershed this was a 59th percentile event. So, while this was extreme for two stations it was not extreme for the subbasin on average.
3. When examining the average percentile hours, we noticed that some historically large events that we expected to fill out the top 10 and top 25 hourly precipitation percentiles were not there, namely the October 24-25, 2021, event. We determined that this was a

result of stations reporting no measured precipitation and therefore dragging down the subbasin station percentile average.

4. Due to this observation, we elect to do an average across the subbasin that did not include zero percentile ranks in the average. This effectively increased the number of events that showed up in higher average percentile ranks.
5. Following our averaging we elected to look at the top five percent of hour precipitation observations (averaged across subbasin 95th percentile and above). When combining the hourly lists from each subbasin, there were 155 hourly observations using the averaging method that included zero percentiles and 352 hourly observations with the averaging method that did not include zero percentiles. When using the same methodology for event definition as above in the NCFR analysis, 84 unique events were identified using the hourly observations that used the non-zero averaging method

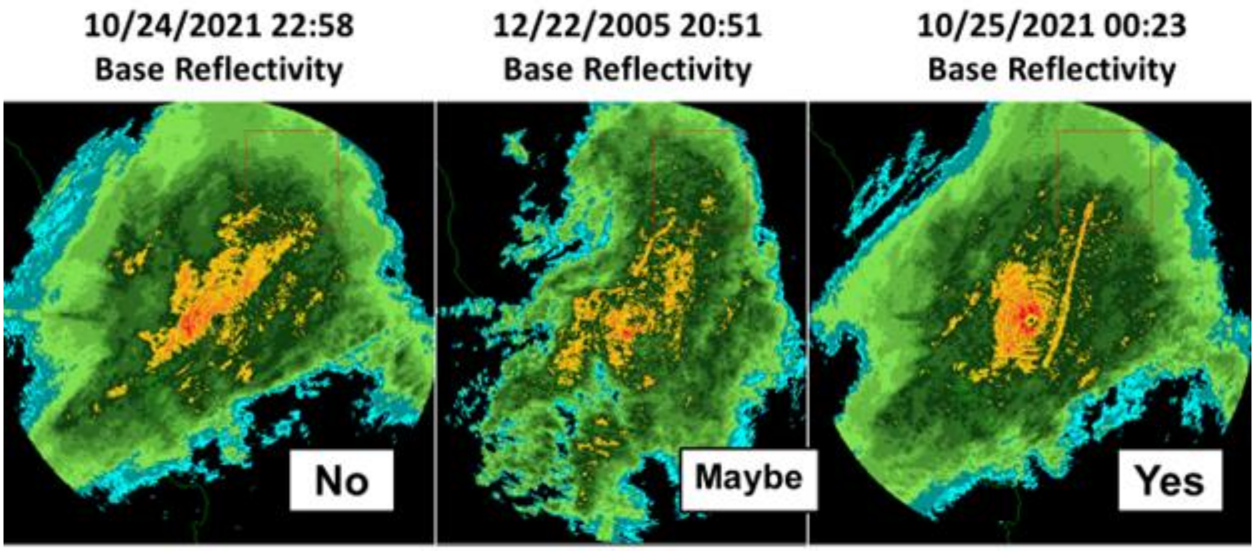


Figure B-28: Example of events that fall under the rubric of no, maybe, and yes for an NCFR over the Yuba and Feather Watersheds via the KDAX (Davis) NEXRAD Radar.

B.3. AR Reconnaissance

Contributing Authors: Anna Wilson, Center for Western Weather and Water Extremes and Minghua Zheng, Center for Western Weather and Water Extremes

Background

Atmospheric River (AR) Reconnaissance (AR Recon) is a targeted observational campaign that fills critical gaps in conventional data sources over the Northeast Pacific Ocean, primarily in and around ARs (Ralph et al., 2020; Zheng et al., 2021a). The main objective of AR Recon is to improve forecasts of landfalling ARs affecting the U.S. West Coast. AR Recon observations do this by enabling better model representation of the precipitation-causing AR storms, which is an essential foundation to enhance the accuracy of precipitation forecasts. They also enable critical diagnostics of model errors, which is the first step to improving model performance. The program is led by CW3E in partnership with NOAA, and is guided by an international, interagency Steering Committee. Operationally, the program includes Air Force Reserve Command's (AFRC) 53rd Weather Reconnaissance Squadron (53 WRS) WC-130J aircraft generally based in Mather, CA, and one National Oceanic and Atmospheric Administration (NOAA) Aircraft Operations Center (AOC) G-IV aircraft generally based in Honolulu, HI.

Using data from AR Recon campaigns back to 2016, studies have demonstrated the positive impact of dropsonde data on AR forecast skill (e.g., Stone et al. 2020; Zheng et al. 2021b and 2023; Lord et al. 2023a,b; DeHaan et al. 2023); more details are provided in Section 3. Alongside the improvement of real-time NWP forecasts, the vast observational datasets collected as part of AR Recon enabled detailed process studies that further the understanding of the key physical processes and dynamics of ARs (e.g., Cannon et al. 2020; Norris et al. 2020; Cobb et al. 2021a; Lavers et al. 2023; Ralph et al. 2023). Observations of ARs can also be used in model assessment studies, such as examining model biases and forecast model skill (e.g., Lavers et al. 2018, 2020a; Stone et al. 2020), and their fidelity compared to reanalysis products (e.g., Guan et al. 2018; Cobb et al. 2021b). These studies show that AR Recon is successful at addressing both operational, real-time forecasting needs, as well as longer-term research goals, within the framework of a Research and Operations Partnership.

In January 2023, the 53 WRS and NOAA AOC flew the longest sequence on record to sample a family of ARs impacting California (DeFlorio et al., 2023), with IOPs for 13 consecutive days, including eight days with multiple aircraft sampling ARs. Airborne Radio Occultation (ARO; Haase et al., 2021) was available on the G-IV and equipment has since been added to all WC-130Js. Additionally, drifting buoys that are part of the NOAA-funded, Scripps Lagrangian Drifter Laboratory-led Global Drifter Program (Centurioni et al., 2017), upgraded by AR Recon to measure surface pressure, were deployed via AFRC 53 WRS and ships of opportunity. Radiosondes were launched throughout California, which is a component of AR Recon funded through USACE FIRO.

This past season represented the highest number of IOPs flown to date (Table B-2, Figure B-29), and the highest number of radiosondes released in a single season (Table B-2, Figure B-30). All AR Recon dropsonde, radiosonde, and buoy data were distributed in real time via the Global Telecommunications System for operational NWP assimilation. Data were assimilated in real-time into GFS, ECMWF IFS, and NAVGEM operational forecast models, amongst others.

	2023	2022	2021	2020	2016-2019 (non-operational)
Number of IOPs	39	25	29.5	17	15

Number of flights	51	32	45	31	19
Dropsondes	1380	746	1142	738	631
Radiosondes	400	11	111	58	259
Drifters Deployed	50	50	30	64	32
ARO profiles	1212	530	872	686	168

Table B-2. *Summary of the number of AR Recon observations taken by year. AR Recon was considered an operational requirement beginning in 2020 (ICAMS, 2022).*

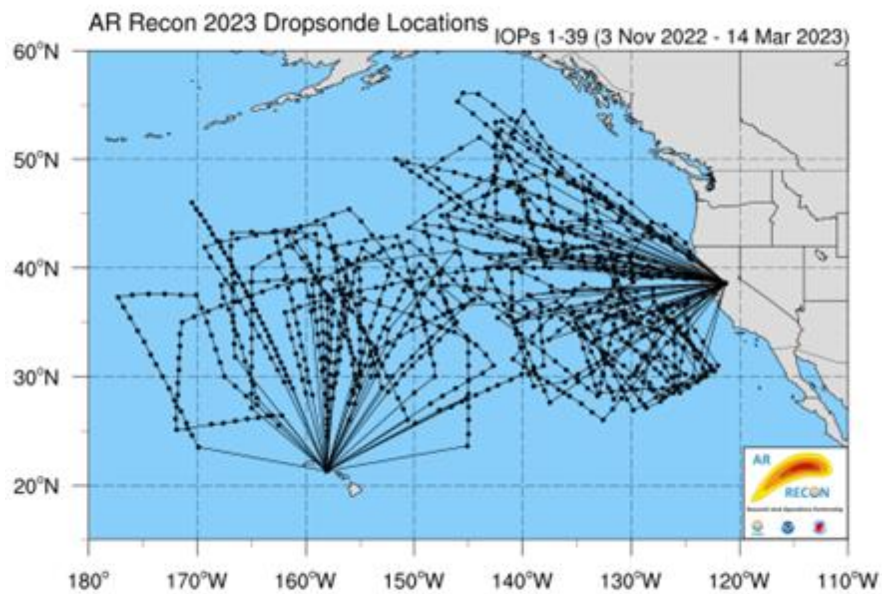


Figure B-29. *AR Recon WY2023 Pacific dropsonde release locations.*

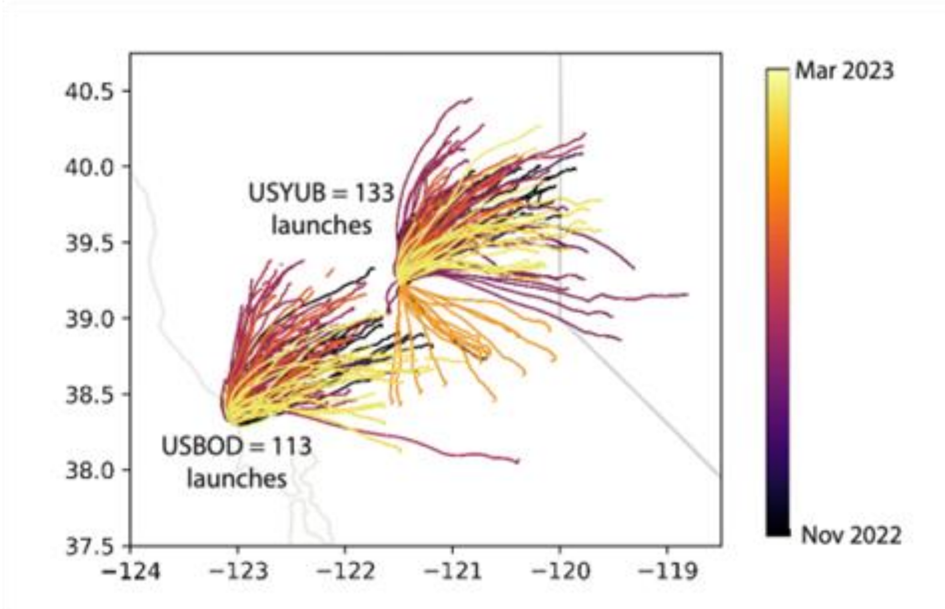


Figure B-30. *Radiosonde release locations in northern California used during AR Recon 2023 with trajectories. USBOD: U.S. Bodega Bay; USYUB: U.S. Yuba.*

AR Recon Impacts

Near real-time data impact experiments with the National Centers for Environmental Prediction operational global forecast system (GFS) have become a regular and important part of the yearly AR Recon campaign, to examine and document the dropsonde impacts on predicting landfalling ARs and their associated precipitation. Hindcast studies from multiple modeling centers have also explored data impact in post-season analyses. Some highlights are below:

- For the GFS model, preliminary results for consecutive IOPs in 2023, 6 (January 6) – 18 (January 18) show large positive impacts on precipitation forecasts over California. The Mean Absolute Error (MAE) improvements in experiments including AR Recon dropsonde data compared to model runs without the data are about 7%, 11%, 25% and 18% for precipitation thresholds of 0.1, 0.5, 1.0 and 2.5 inches, respectively. The domain average MAE improvement in runs including the dropsonde data is nearly 18%. Dropsondes were also found to be capable of reducing cold, dry, and slow biases in GFS model background forecasts for AR conditions.
- Preliminary results from case studies with the GFS model in Washington, the central coast in CA, and the Sierras have all illustrated increased lead times on the order of days for the regions of heaviest precipitation, with inclusion of dropsondes compared to not including the dropsondes.
- AR Recon per-observation impact on forecasts is more than double that of the individual observations in North American Radiosonde network, with global reduction in forecast

error per flight comparable to the entire North American Radiosonde network in the Naval Research Laboratory (NRL)'s global model (Stone et al., 2020).

- AR Recon data improve precipitation forecasts up to 9% more than satellite data alone, particularly in heavy rainfall, in a National Center for Atmospheric Research-led study (Sun et al 2022).
- AR Recon dropsonde observations reduce errors in AR water vapor flux and inland precipitation at forecast lead times from 1 to 6 days, with the largest improvements in inland precipitation forecast skill associated with back-to-back flights every other day (Zheng et al 2021b). There was also a case study in this paper that showed the precipitation forecast error over western Washington was improved by ~50% at a lead time of 12–36 h.
- In the GFS, at critical lead times of 3-5 days, AR Recon observations improve precipitation forecasts by 5-15% over the entire western US and by 10-20% over Pacific Northwest and Northern California (Lord et al 2023a).
- In the GFS, AR Recon observations improve atmospheric river, wind, humidity, and precipitation forecasts over the U.S. West Coast by improving low-altitude moisture fields. (Lord et al 2023b)
- Assimilating the AR Recon drifter observations showed beneficial impact on Northern Hemisphere and North American forecasts in the NRL model such as improving 72-h and 96-h Northern Hemisphere forecasts of winds in the lower and middle troposphere, and geopotential height in the whole troposphere (Reynolds et al., 2023).
- AR Recon observations increase the assimilation number of satellite data by 5-10% into the NCEP Global Forecast System, reduces bias correction, and benefits the data assimilation processes for up to 1 week after each mission (Zheng et al., 2022).
- Dropsondes improved the representation of ARs in the model analyses, particularly near sharp horizontal and vertical gradients. Reduced mission frequency and dropsonde horizontal spacing degraded forecast skill (Zheng et al., 2023).
- The ECMWF and GFS models show many improvements in forecast skill with added information from dropsondes, w/ significant improvements in the forecast IVT and precipitation generally occurring in both models. Several case studies conducted looking at FIRO watersheds in particular illustrated improvements in the Yuba-Feather system (Figure B-31) (DeHaan et al., 2023).

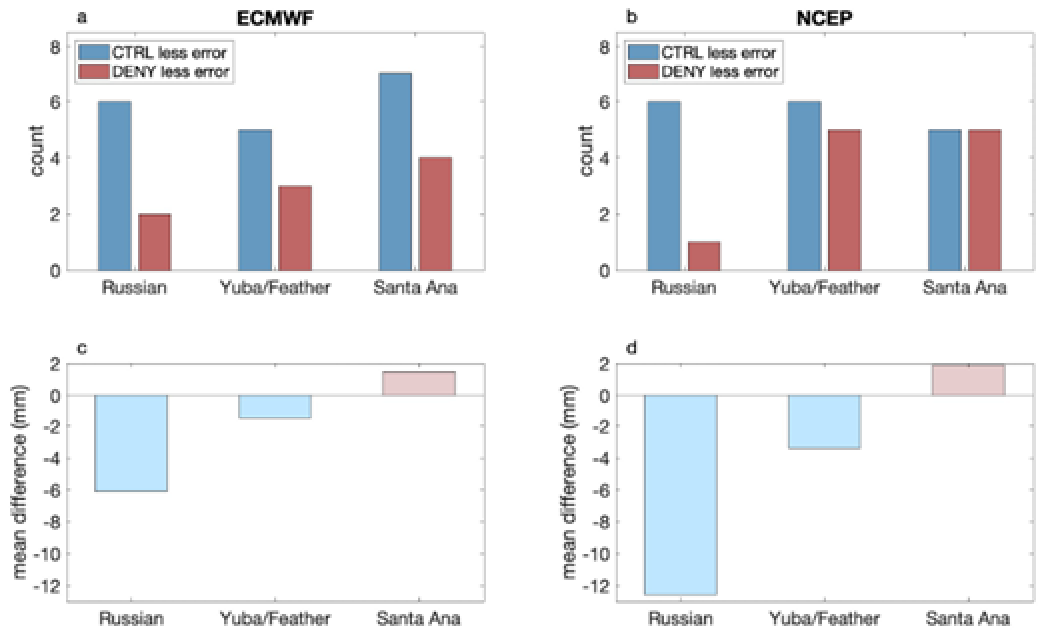


Figure B-31. *(from DeHaan et al., 2023): Counts of instances (valid days and lead times) where each of the control and denial forecasts had a smaller watershed intensity error magnitude for 3 California watersheds (a and b), and the mean difference (control - denial) of the magnitude of the error for those instances (c and d). The counts are limited to cases where the difference between control and denial is greater than 1mm/24h.*

Recommendations

- Continue AR Recon field campaigns yearly, including exploring innovative new technology as well as collecting observations we understand are critical.
    - Continue appropriate advances, including data collection farther upstream.
    - Continue to conduct near real time data denials and add near real time model verification
- Continue refining assessment protocols useful for the Yuba Feather FIRO program, within the Research and Operations Partnership framework.
    - Assess data impacts in the Yuba-Feather in the framework of probabilistic forecasts.
- Begin data assimilation studies for radiosonde data.
- Continue to conduct case studies designed to investigate dynamical and physical mechanisms behind forecast improvements.

References

Cannon, F., N.S. Oakley, C.W. Hecht, A. Michaelis, J.M. Cordeira, B. Kawzenuk, R. Demirdjian, R. Weihs, M.A. Fish, A.M. Wilson, and F.M. Ralph, 2020: Observations and Predictability of a High-Impact Narrow Cold-Frontal Rainband over Southern California on 2 February 2019. *Wea. Forecasting*, **35**, 2083–2097.

Centurioni, L., A. Horaìnyi, C. Cardinali, E. Charpentier, and R. Lumpkin, 2017: A global ocean observing system for measuring sea level atmospheric pressure: Effects and impacts on numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **98**, 231–238.

Cobb, A., A. Michaelis, S. Iacobellis, F.M. Ralph, and L. Delle Monache, 2021a: Atmospheric river sectors: Definition and characteristics observed using dropsondes from 2014-2020 CalWater and AR Recon. *Mon. Wea. Rev.*, **149**, 623-644.

Cobb, A., Delle Monache, L., Cannon, F., & Ralph, F. M., 2021b: Representation of dropsonde- observed atmospheric river conditions in reanalysis. *Geophys. Res. Lett.*, **48**, e2021GL093357.

DeFlorio, M.J., A. Sengupta, C.M. Castellano, J. Wang, Z. Zhang, A. Gershunov, K. Guirguis, R. Luna Niño, R.E. Clemesha, M. Pan, M. Xiao, B. Kawzenuk, P.B. Gibson, W. Scheftic, P.D. Broxton, M.B. Switanek, J. Yuan, M.D. Dettinger, C.W. Hecht, D.R. Cayan, B.D. Cornuelle, A.J. Miller, J. Kalansky, L. Delle Monache, F.M. Ralph, D.E. Waliser, A.W. Robertson, X. Zeng, D.G. DeWitt, J. Jones, and M.L. Anderson: From California's extreme drought to major flooding: Evaluating and synthesizing experimental seasonal and subseasonal forecasts of landfalling atmospheric rivers and extreme precipitation during Winter 2022 - 2023. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-22-0208.1.

DeHaan, L.L., Wilson, A.M., Kawzenuk, B., Zheng, M., Delle Monache, Wu, X., Lavers, D.A., Ingleby, B., Tallapragada, V., Pappenberger, F., Ralph, F.M., 2023: Impacts of Dropsonde Observations on Forecasts of Atmospheric Rivers and Associated Precipitation in the NCEP GFS and ECMWF IFS models. *Mon. Wea. Rev.*, https://doi.org/10.1175/WAF-D-23-0025.1.

Guan, B., D. Waliser, and F. Ralph, 2018: An inter-comparison between reanalysis and dropsonde observations of the total water vapor transport in individual atmospheric rivers. *J. Hydrometeor.*, **19**, 321-337.

Haase, J. S., Murphy, M. J., Cao, B., Ralph, F. M., Zheng, M., & Delle Monache, L., 2021:

Multi-GNSS airborne radio occultation observations as a complement to dropsondes in atmospheric river reconnaissance. *J. Geophys. Res.: Atmos.* **126**, e2021JD034865.

Interagency Meteorology Coordination Office (ICAMS), 2022: The 2022 National Winter Season Operations Plan. Available at https://www.icams-portal.gov/resources/ofcm/nwsop/2022_nwsop.pdf

Lavers, D.A., M.J. Rodwell, D.S. Richardson, F.M. Ralph, J.D. Doyle, C.A. Reynolds, V. Tallapragada, and F. Pappenberger, 2018: The Gauging and Modeling of Rivers in the Sky. *Geophys. Res. Lett.*, **45**, 7828–7834, doi:10.1029/2018GL079019

Lavers, D.A., N.B. Ingleby, A.C. Subramanian, D.S. Richardson, F.M. Ralph, J.D. Doyle, C.A. Reynolds, R.D. Torn, M.J. Rodwell, V. Tallapragada, and F. Pappenberger, 2020: Forecast Errors and Uncertainties in Atmospheric Rivers. *Wea. Forecasting*, **35**, 1447–1458, https://doi.org/10.1175/WAF-D-20-0049.1

Lavers, D. A., Torn, R. D., Davis, C., Richardson, D. S., Ralph, F. M., & Pappenberger, F., 2023: Forecast Evaluation of the North Pacific Jet Stream Using AR Recon Dropwindsondes, *Quart. J. Roy. Meteor. Soc*., https://doi.org/10.1002/qj.4545.

Lord, S. J., Wu, X., Tallapragada, V. and Ralph, F.M., 2023a: The Impact of Dropsonde Data on the Performance of the NCEP Global Forecast System During the 2020 Atmospheric Rivers Observing Campaign. Part 1: Precipitation. *Wea. Forecasting*, **38(1)**, 17-45.

Lord, S. J., S. J., Wu, X., Tallapragada, V., & Ralph, F.M., 2023b: The Impact of Dropsonde Data on the Performance of the NCEP Global Forecast System During the 2020 Atmospheric Rivers Observing Campaign. Part II: Dynamic Variables and Humidity, *Wea. Forecasting*, **38(5)**, 721-752.

Norris, J.R., F.M. Ralph, R. Demirdjian, F. Cannon, B. Blomquist, C.W. Fairall, J.R. Spackman,

S. Tanelli, and D.E. Waliser, 2020: The Observed Water Vapor Budget in an Atmospheric River over the Northeast Pacific. *J. Hydrometeor.*, **21**, 2655–2673.

Ralph, F. M., M. Simpson, S. Iacobellis, J. M. Cordeira, F. Cannon, A. C. Michaelis, L. Delle Monache, and A. Cobb, 2023: Dropsonde observations of the stable marine boundary layer beneath atmospheric rivers. *Mon. Wea. Rev*., submitted.

Ralph, F. M., F. Cannon, V. Tallapragada, C. A. Davis, J. D. Doyle, F. Pappenberger, A. Subramanian, A. M. Wilson, D. A. Lavers, C. A. Reynolds, J. Haase, J. J. Rutz, J. M.Cordeira, M. Zheng, C. W. Hecht, B. Kawzenuk, L. Delle Monache, 2020: West Coast Forecast Challenges and Development of Atmospheric River Reconnaissance. *Bull. Amer. Meteor. Soc.*, **101**, E1357–E1377, https://doi.org/10.1175/BAMS-D-19-0183.1.

Reynolds, C. A., Stone, R. E., Doyle, J. D., Baker, N. L., Wilson, A. M., Ralph, F. M., Lavers, D. A., Subramanian, A. C., & Centurioni, L., 2023: Impacts of Northeastern Pacific Buoy Surface Pressure Observations. *Mon. Wea. Rev.*, **151(1)**, 211-226.

Stone, R.E., Reynolds, C.A., Doyle, J.D., Langland, R.H., Baker, N.L., Lavers, D.A. and Ralph, F.M., 2020. Atmospheric river reconnaissance observation impact in the Navy global forecast system. *Mon. Wea. Rev.*, **148(2)**, 763-782.

Sun, W., Liu, Z., Davis, C. A., Ralph, F. M., Delle Monache, L., Zheng, M., 2022. Impacts of dropsonde and satellite observations on the forecasts of two atmospheric-river-related heavy rainfall events. *Atmos. Res.,* **278**, 106327.

Zheng, M., L. Delle Monache, X. Wu, F.M. Ralph, B. Cornuelle, V. Tallapragada, J.S. Haase,

A.M. Wilson, M. Mazloff, A. Subramanian, and F. Cannon, 2021a: Data Gaps within Atmospheric Rivers over the Northeastern Pacific. *Bull. Amer. Meteor. Soc.*, **102**, E492–E524, https://doi.org/10.1175/BAMS-D-19-0287.1

Zheng, M., L. Delle Monache, B.D. Cornuelle, F.M. Ralph, V.S. Tallapragada, A. Subramanian,

J.S. Haase, Z. Zhang, X. Wu, M.J. Murphy, T.B. Higgins, and L. DeHaan, 2021b: Improved Forecast Skill through the Assimilation of Dropsonde Observations from the Atmospheric River Reconnaissance Program. *J. Geophys. Res.-Atmos.*, **126**, e2021JD034967.

Zheng, M., Delle Monache, L., Wu, X., Kawzenuk, B., Ralph, F.M., Zhu, Y., Torn, R.,

Tallapragada, V., and others, 2022. Impact of Atmospheric River Reconnaissance Dropsonde Data on the Assimilation of Satellite Data in GFS. *Geophys. Res. Lett.*, DOI: 10.1002/essoar.10510741.1

Zheng, M., Torn, R., Delle Monache, L., Doyle, J., Ralph, F.M., Tallapragada, V., Davis, C., Steinhoff, D., Wu, X., Wilson, A., Papadopoulos, C., Mulrooney, P., 2023: Improved Sampling Strategies for Atmospheric River Reconnaissance. *Mon. Wea. Rev.*, accepted pending revision.

B.4. Lead-time prediction of landfalling ARs

Contributing Authors: Jason M. Cordeira, Center for Western Weather and Water Extremes

Introduction

Atmospheric Rivers (ARs) are long and narrow regions of enhanced integrated water vapor transport (IVT) that can influence the occurrence of precipitation-related high-impact weather events along U.S. West Coast such as floods and flash floods (e.g., Young et al. 2017). Landfalling ARs may also influence the occurrence of extreme wind events (Waliser and Guan 2017) and increase the likelihoods of avalanches and avalanche fatalities (Hatchett et al. 2017) and shallow landslides (Oakley et al. 2018).

The potential for hazardous weather associated with landfalling ARs can be summarized by (1) National Weather Service-issued watches, warnings, and advisories (WWAs) where 60–90% of flood-related WWAs in the Western U.S. occur on days with cool-season landfalling ARs (Bartlett and Cordeira 2021) and (2) damage claims in the National Flood Insurance Program where ARs have caused an average of $1.1 billion in flood damages annually across the Western U.S. (Corringham et al. 2019).

Due to the causal relationship between landfalling ARs and the potential for hazardous weather across the western US, reliable and skillful forecasts of landfalling ARs are critical to hazard preparation, risk mitigation, and water resources management (e.g., DeFlorio et al. 2018; Ralph et al. 2020; Cordeira and Ralph 2021).

The goal of this study is to objectively evaluate the lead-time prediction skill of the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) model forecasts of enhanced IVT in northern California for October 2017 through January 2023 that is observed during landfalling ARs. This study complements previous studies Cordeira and Ralph (2021) and Stewart et al (2022).

Methodology and Analysis

This study leverages forecast data that are commonly illustrate via the "CW3E" AR Landfall Tool. To summarize forecasts of AR frequency, intensity, duration, and landfalling location, an "AR Landfall Tool" was created and used as an aid in providing situational awareness of landfalling ARs using ensemble numerical weather prediction data (Cordeira et al. 2017; Cordeira and Ralph 2021).

The AR Landfall Tool is primarily a tool that depicts ensemble IVT data as a probability-over-threshold for different IVT magnitudes for a forecast transect along the west coast of North America. The most used threshold in AR-related forecasts along the U.S. West Coast is IVT magnitudes ≥250 kg m^{-1} s^{-1} (Cordeira et al. 2017) and the ensemble probability of IVT magnitudes ≥250 kg m^{-1} s^{-1} is referred to as P_{250} by Cordeira and Ralph (2021).

The study by Cordeira and Ralph (2021) found that the 20-member GEFS P_{250} forecasts near coastal northern California at 38ºN, 123ºW for WY2017–2020 were reliable and successful at lead times of ~8–9 days with an average success ratio >0.50 for P_{250} forecasts ≥50% at lead times of 8 days and Brier skill scores >0.10 at a lead time of 8–9 days. The highest success ratios and probability of detection values for P_{250} forecasts ≥50% occurred on average along the northern California and Oregon coastlines and the lowest occurred on average along the southern California coastline. The average probability of detection of more intense and longer duration landfalling ARs was also 0.10–0.20 higher than weaker and shorter duration events at lead times of 3–9 days.

In October 2021, the NCEP GEFS was upgraded to include 30 ensemble members and the AR Landfall Tool has been continuously run with archived data through WY2023. As opposed to visualizing individual forecasts from within the AR Landfall Tool framework along the West Coast, a different forecast visualization can be constructed to show how individual forecasts evolve as a function of lead time as in Fig. 4 of Stewart et al. (2022). These "waterfall analyses" allow for a visual assessment of the lead-time prediction (i.e., situational awareness) of enhanced odds of a landfalling AR at a given point. A waterfall analysis for the verification period from 17 December 2022 through 17 January 2023 is shown in Figure B-32 at 37.5N, 122.5W near San Francisco, California.

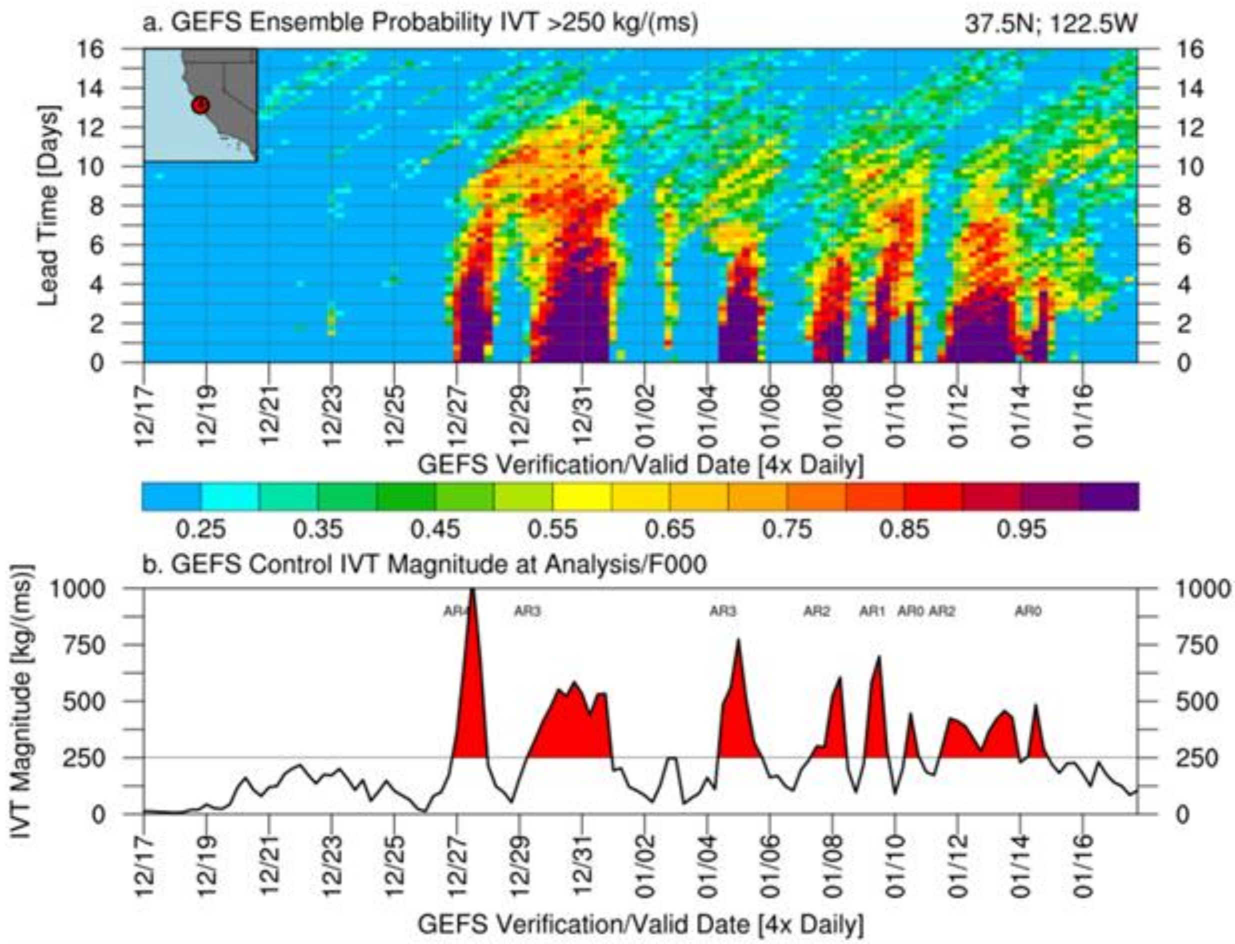


Figure B-32. *(a) A verification time – lead-time analysis of the ensemble odds of IVT magnitudes >=250 kg/ms (shaded) from the NCEP GEFS for forecasts verifying between 17 December 2022 through 17 January 2023. (b) A time series of IVT magnitude from the GEFS control IVT magnitude at the 0-hour forecast time (kg/ms) with periods with IVT magnitudes >=250 kg/ms shaded in red.*

While the waterfall analysis in Figure B-32 can be subjectively analyzed to identify periods of false alarms (e.g., 29 December 2022; 3 January 2023), periods with longer-lead prediction (e.g., 31 December 2022), or periods with shorter-lead prediction (e.g., 8 January 2023), it does not allow for an objective assessment of "how far in advance does the GEFS provide enhanced situational awareness for a landfalling AR?" To answer this question, we investigate the lead time at which the probability-over-threshold increases above a given percentage and stays above a given threshold for the subsequent leads prior to verification. This lead time value can then be averaged for the entire period of AR landfall (i.e., its duration), taken at the start time of the landfalling AR, or the time of maximum IVT magnitude to answer slightly different questions about the afforded situational awareness.

Northern California using IVT information at 37.5N, 122.5W (near SFO)								
	Start	End	Max IVT (kg m^{-1} s^{-1})	Duration (h)	AR Scale	Avg IVT Dir (deg)	tIVT (10^7 kg m^{-1})	Avg Z0C (m ASL)
AR #1	2022-12-26 2300 UTC	2022-12-27 2300 UTC	1095.8	25	4	251	6.01	3342
AR #2 & #3	2022-12-29 0700 UTC	2022-12-31 2100 UTC	606.1	63	3	266	10.49	3071
AR #4	2023-01-04 1000 UTC	2023-01-05 1800 UTC	832.1	33	3	195	6.00	2397
AR #5	2023-01-07 1000 UTC	2023-01-08 1200 UTC	664.7	27	2	202	3.76	2213
AR #6	2023-01-09 0300 UTC	2023-01-09 1800 UTC	817.5	16	2	215	3.29	2582
AR #6	2023-01-10 0700 UTC	2023-01-10 1700 UTC	618.9	11	1	215	1.38	1904
AR #7	2023-01-11 1100 UTC	2023-01-13 2300 UTC	598.9	61	3	198	7.96	2933
AR #8	2023-01-14 0600 UTC	2023-01-14 1800 UTC	601.6	13	1	213	1.82	2213

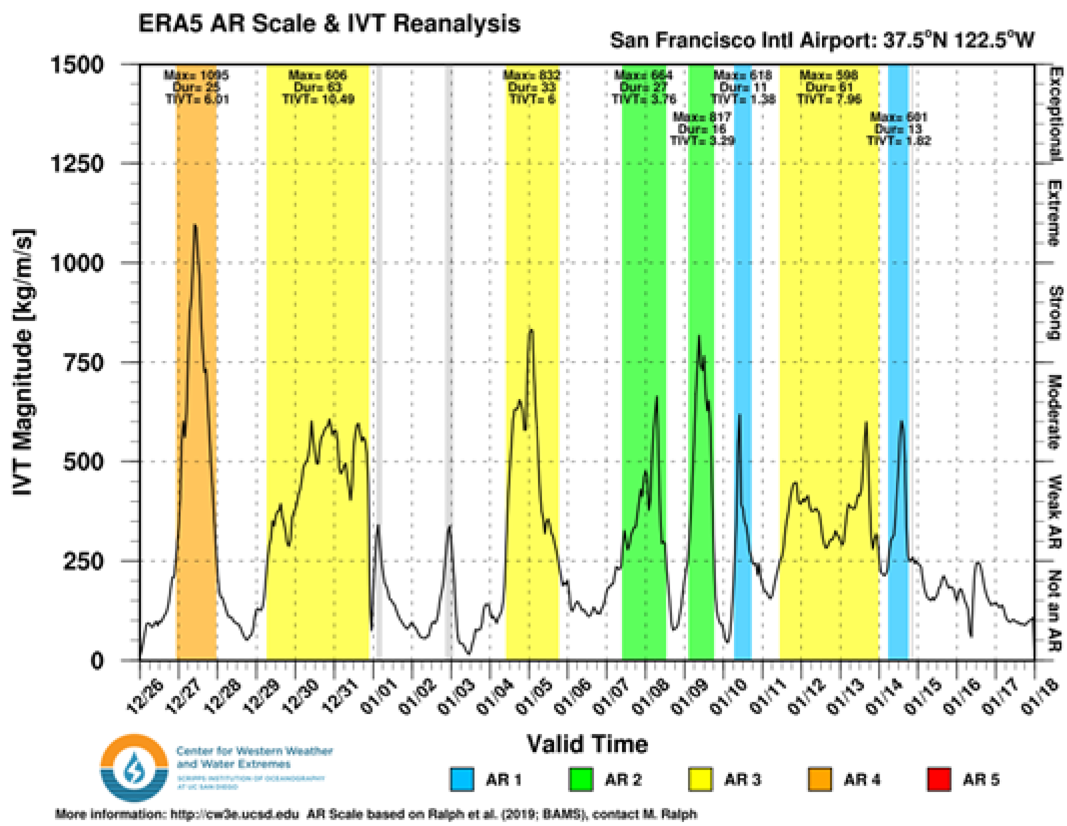


Figure B-33. *Table of the eight landfalling AR events, their characteristics, and IVT magnitude at 37.5N, 122.5W during the deep dive period from late December 2022 through January 2023.*

The above-mentioned analysis of the lead-time of probability-over-threshold will be assessed for the eight landfalling ARs and use an initial threshold for the probability-over-threshold increasing above 75% and subsequently staying above a more generous 50% prior to verification (Fig. 3). For each of the landfalling ARs, the lead time prediction for the time of maximum IVT magnitude and the event average duration was greater than the start time of the AR (i.e., the start time of an AR is more difficult to predict than a time in the middle of AR landfall or its average over its duration). The average lead time for the start time of the AR was ~3 days with a range from 1-4 days, whereas the average lead time for either the event or time of maximum IVT was ~5 days with a range from ~9-10 days to ~4 days.

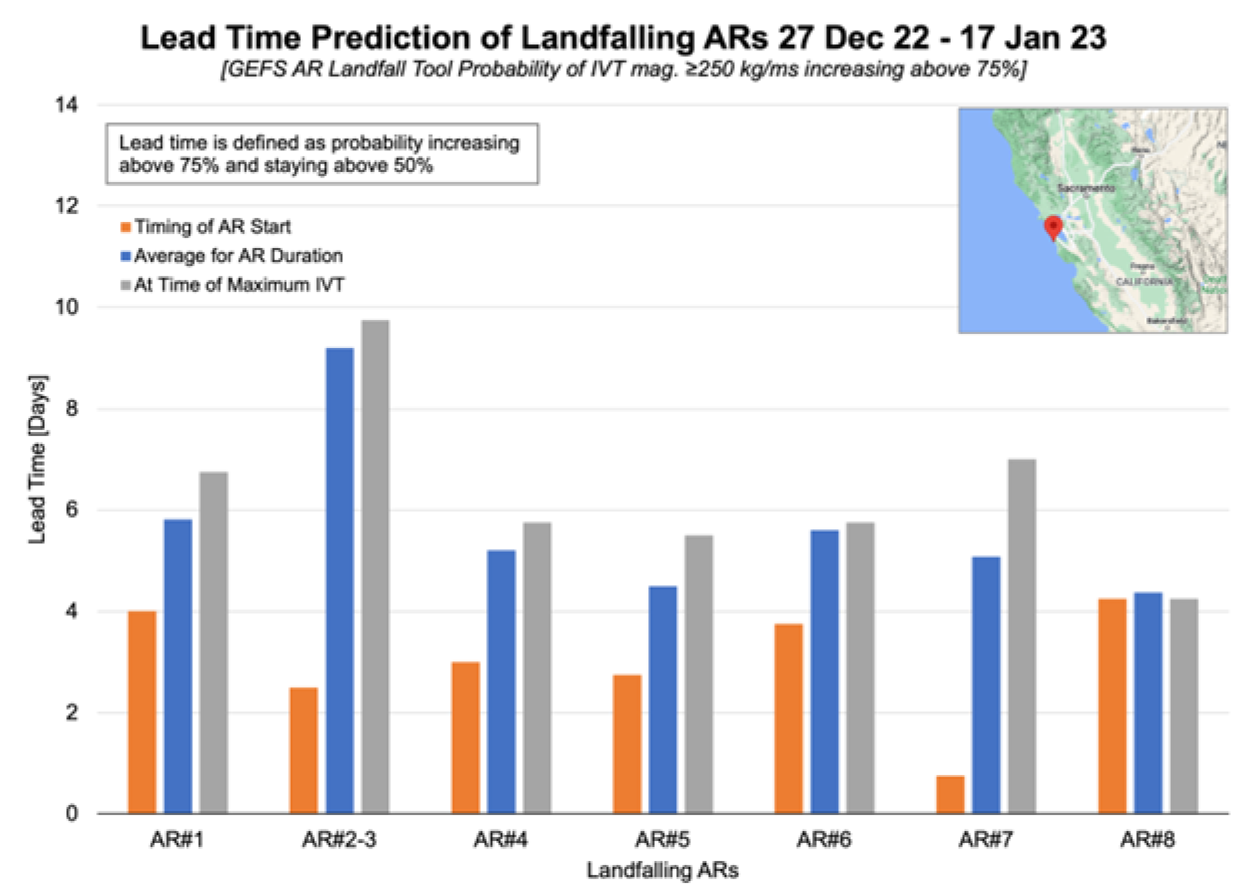


Figure B-34. *Lead time at which the ensemble probability-over-threshold increased above 75% and stayed above 50% for the start time of the AR (orange), the time of maximum IVT magnitude during the AR (gray), and an average for all times with the AR (blue) at 37.5N.*

For a longer period of record, we can assess the lead time of the probability-over-threshold for all storms for the period from October 2016 through January 2023 with an AR rank of AR2+ following the AR scale of Ralph et al. (2019). In this analysis, we will use an initial threshold for the probability-over-threshold increasing above 50%, 66%, 75%, or 90% and subsequently staying above those same values prior to verification (Figure B-35).

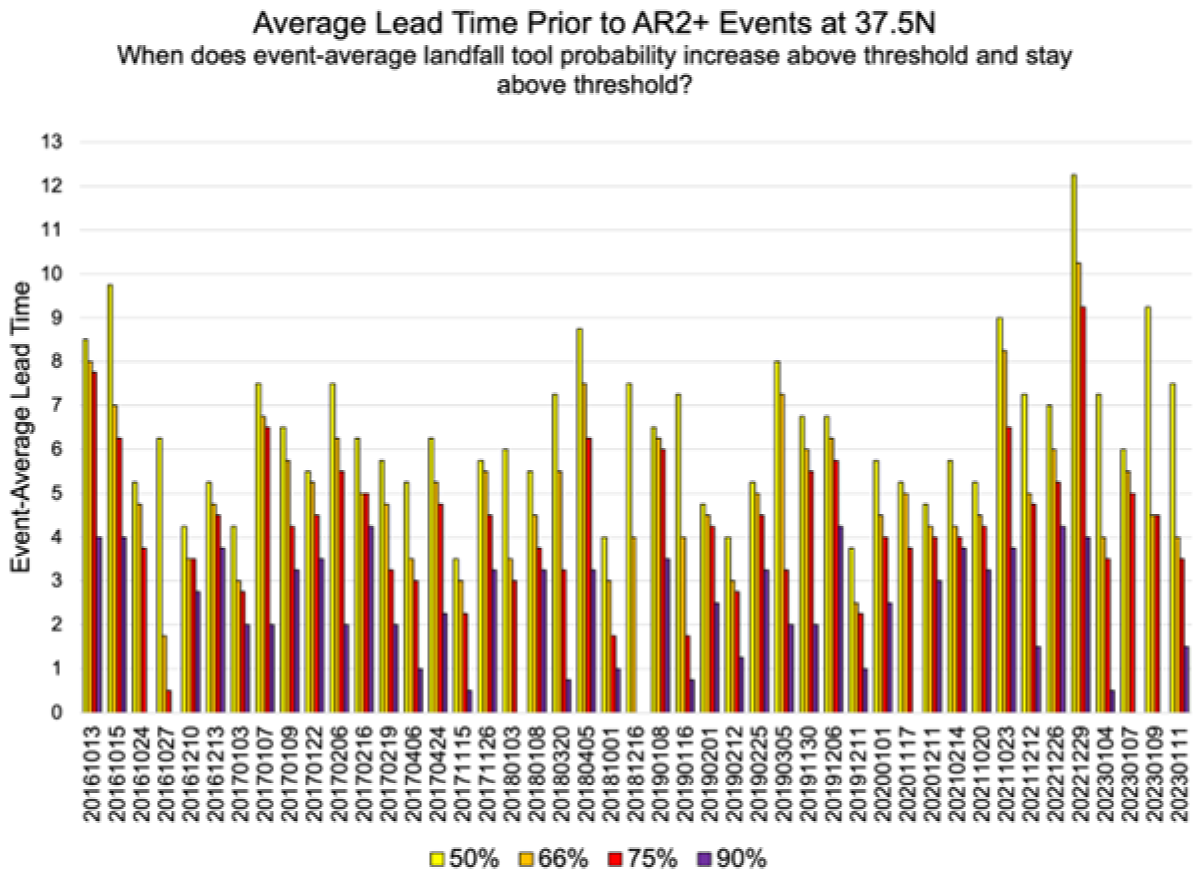


Figure B-35. Lead time at which the ensemble probability-over-threshold increased above 50%, 66%, 75%, or 90% and stayed above those percentages for the average of all times within an AR2+ event at 37.5N.

In this analysis, the late December 2022 event was predicted with the longest lead-time prediction using this metric as compared to any prior landfalling AR2+ event since October 2016 with 66% and 75% odds appearing (and staying above) at lead times >9 days, and with 50% odds appearing >12 days in advance. The average values across all storms during this period of record include 6.4 days (50%), 5.0 days (66%), 4.2 days (75%), and 2.2 days (90%).

Summary

The framework developed by the CW3E AR Landfall Tool is leveraged to investigate the lead-time predictability or situational awareness of the probability-over-threshold values for a given location. This analysis focuses on landfalling ARs at 37.5N near San Francisco to investigate the lead-time prediction for events during the “Deep-Dive” period for December 2022-January 2023 and again for all AR2+ storms for a period for October 2016 thorugh January 2023. In both analyses it was shown that the NCEP GEFS ensemble probability-over-threshold can provide situational awareness and information of an impending storm at lead times of ~5 days using a 66% threshold with individual events such as the one in late December 2022 containing lead times >9 days, or individual events such as the one in February 2019 containing lead times <3 days.

References

Bartlett, S. M., and J. M. Cordeira, 2021: A climatology study of National Weather Service watches, warnings, and advisories and landfalling atmospheric rivers in the Western U.S. 2006–2018, *Wea. Forecasting,* **36,** 1097–1112.

Cordeira, J. M., and F. M. Ralph, 2021: A Summary of GFS Ensemble Integrated Water Vapor Transport Forecasts and Skill Along the U.S. West Coast during Water Years 2017–2020. *Wea. Forecasting,* **36**, 361–377.

Cordeira, J. M., F. M. Ralph, A. Martin, N. Gaggini, J. R. Spackman, P. J. Neiman, J. J. Rutz, and R. Pierce, 2017: Forecasting atmospheric rivers during CalWater 2015. *Bull. Amer. Meteor. Soc.*, **98**, 449–459.

Corringham, T. W., F. M. Ralph, A. Gershunov, D. R. Cayan, and C. A. Talbot, 2019: Atmospheric rivers drive flood damages in the western United States. *Sci. Adv.*, **5**, 1–7.

Hatchett, B. J., S. Burak, J. J. Rutz, N. S. Oakley, E. H. Bair, and M. L. Kaplan, 2017: Avalanche fatalities during atmospheric river events in the Western United States. *J. Hydrometeor.*, **18**, 1359–1374.

Oakley, N. S., J. T. Lancaster, B. J. Hatchett, J. Stock, F. M. Ralph, S. Roj, and S. Lukashov, 2018: A 22-year climatology of cool season hourly precipitation thresholds conducive to shallow landslides in California. *Earth Interactions*, **22**, 1–35.

Stewart, B. E., Cordeira, J. M., & Ralph, F. M. (2022). Evaluating GFS and ECMWF Ensemble Forecasts of Integrated Water Vapor Transport Along the U.S. West Coast, *Wea. and Forecasting,* **37,** 1985–2004.

Waliser, D., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nature Geoscience*, **10**, 179–183.

Young, A. M., K. T. Skelly, and J. M. Cordeira, 2017: High-impact hydrologic events and atmospheric rivers in California: An investigation using the NCEI Storm Events Database. *Geophys. Res. Lett.*, **44**, 3393–3401.

B.5 West-WRF and its skill: Machine learning and A.I.

Contributing Authors: Agniv Sengupta, Center for Western Weather and Water Extremes

Background

Quantitative precipitation forecasts (QPFs) and forecasts of integrated vapor transport (IVT) provide crucial information to water managers for mitigating urban, riverine, and flash flood risks. In addition, such forecasts have the potential to guide decisions related to reservoir, agricultural, and irrigation management. Forecasts from numerical weather prediction (NWP) models have a prominent role in informing better decision-making. NWP models are based on our best understanding of the dominant physical processes and the most advanced numerical procedures to integrate the equations describing atmospheric evolution. However, they are contaminated by errors in initial conditions, numerical approximations, incomplete understanding of underlying physical processes, and the inherent chaotic nature of the atmosphere.

Recent investigations by the CW3E Machine Learning team have demonstrated that a significant portion of NWP model errors can be recovered in a post-processing framework leveraging recent advancements in artificial intelligence and machine learning. These algorithms can be trained to learn the dynamical model behavior over a historical period when predictions from the same model and observations of the quantity of interest are available.

Machine Learning-based Forecast Products

For the Yuba-Feather FIRO project, the CW3E Machine Learning team has been developing state-of-the-art ML algorithms to improve predictions of extreme weather events, with an emphasis on the prediction (both deterministic and probabilistic) of IVT and precipitation associated with ARs. These algorithms can leverage valuable information provided by West-WRF and learn a significant portion of their biases, resulting in improved forecasts and reliable uncertainty quantification.

1) Development of probabilistic QPFs leveraging the 34-year West-WRF Reforecast:

A deep learning framework based on a U-Net convolutional network (Figure B-36) has been developed (Hu et al. 2023, *MWR*) for post-processing deterministic West-WRF predictions of precipitation and generating 0-5-day probabilistic forecasts of daily accumulated precipitation. This novel deep learning method was tested against state-of-the-art benchmark methods, including an Analog Ensemble, non-homogeneous regression, and mixed-type meta-Gaussian distribution. The U-Net was found to outperform the benchmark methods at all lead times, as measured by Continuous Ranked Probability and Brier skill scores, while producing a reliable estimation of forecast uncertainty, as measured by binned spread-skill relationship diagrams. Additionally, the U-Net was found to have the best performance for extreme precipitation events, i.e., the 95th and 99th percentiles of the distribution.

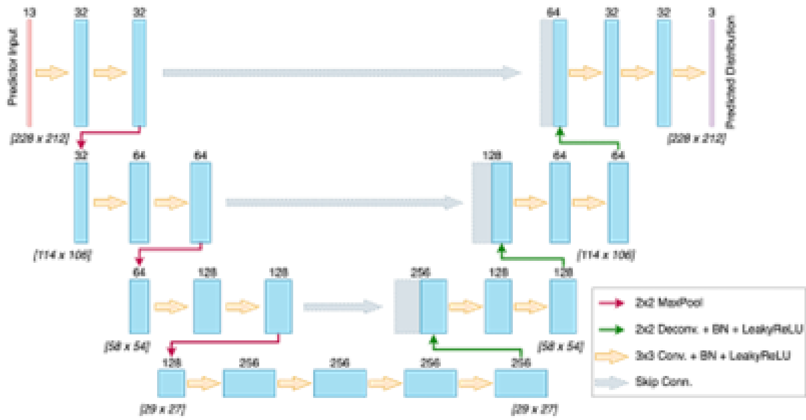


Figure B-36. Designed U-Net model architecture

Figure B-37 shows the mean areal precipitation over the Yuba–Feather watershed, i.e., precipitation aggregated over the wet season starting in December 2016 from the developed U-Net (red) compared to the dynamical benchmark (WRF; blue) and other baseline methods; PRISM dataset is the ground truth (black). The raw West-WRF remained close to PRISM until 18 January 2017 when West-WRF started to overpredict. There were several major precipitation events on 9 January 2017, 19 January 2017, 27 February 2017, and 21 February 2017, and West-WRF showed significant overprediction during the latter three events, enlarging the difference to PRISM. On the other hand, other benchmark methods (MMGD, CSGD, and AnEn) all showed underprediction early on during the water year around 13 December 2016, with MMGD consistently producing the most underprediction among the others. The proposed U-Net model (red line) closely follows PRISM throughout the year and its prediction for the year-round total precipitation is the most accurate compared to other baseline forecasts.

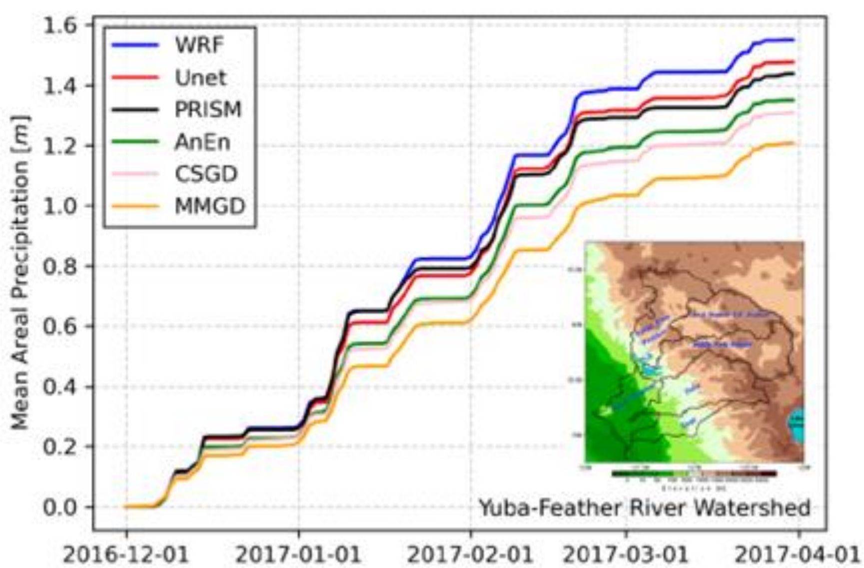


Figure B-37. Mean areal precipitation over the Yuba-Feather watershed in eater year 2017: Proposed U-Net vs West-WRF and other traditional post-processing methods. PRISM is used as the observational ground-truth dataset.

The West-WRF model post-processed with deep learning (West-WRF + Unet) has now been implemented operationally into CW3E's near real-time (NRT) operational forecast system (Figure B-38) and is currently generating probabilistic QPFs and probabilities of 24-hour precipitation exceeding various thresholds (e.g., > 1 mm, > 10 mm, > 25 mm, >50 mm, >100 mm) throughout the water year.

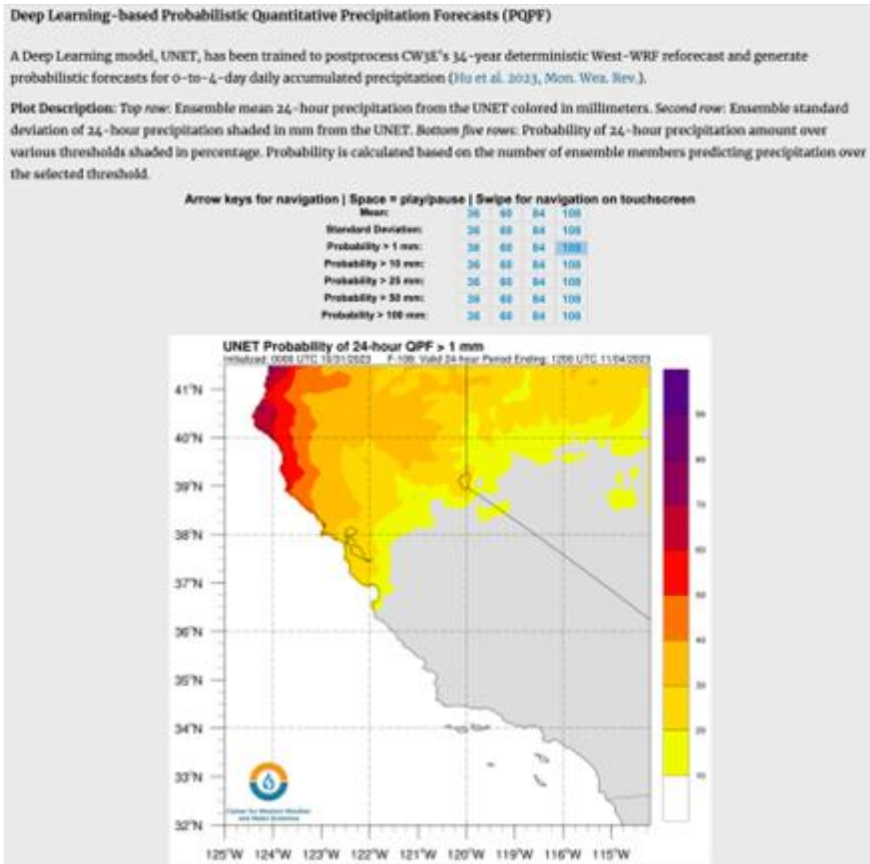


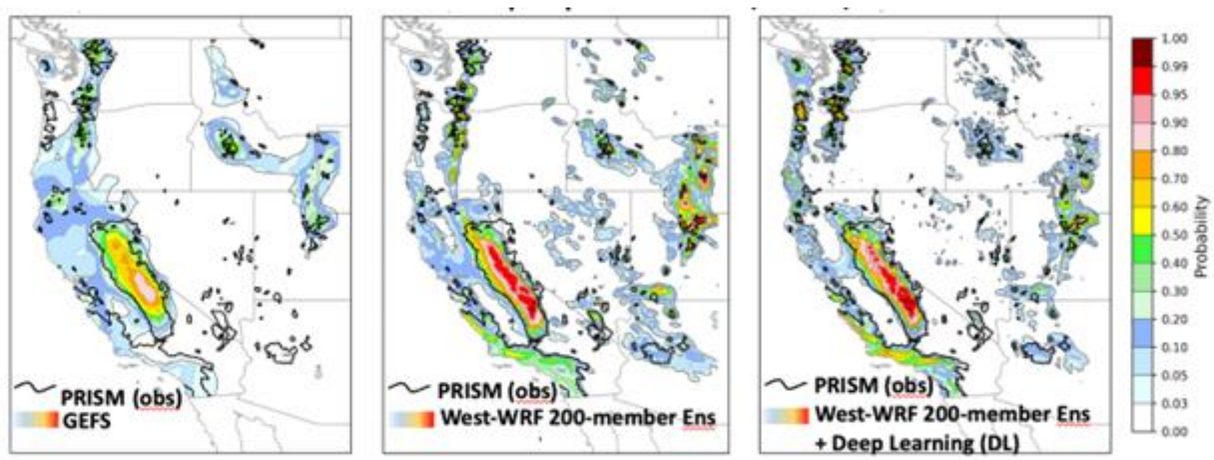
Figure B-38. Associated near real-time (NRT) forecast product generating probabilistic QPFs operationally throughout the water year via CW3E's NRT operational system

2) Application of deep learning on CW3E's West-WRF 200-member NRT ensemble:

For this task, we utilized NWP model outputs from the 200-member ensemble of a customized version of the Weather Research and Forecasting (WRF) model, named West-WRF, developed by CW3E that is run in near-real-time forecast mode at 9-km resolution in support of decision making and scientific research of extreme weather events over the Western U.S. The deep learning application involves an *Artificial Neural Network–Censored, Shifted Gamma Distribution* model (ANN-CSGD; Ghazvinian et al. 2022) for generating post-processed, high-resolution, probabilistic precipitation forecasts for lead times up to 7 days.

Figure B-39 shows the probabilities of 24-hour accumulated precipitation > 1 inch for a representative test case (24 December 2021) from GEFS (Figure B-39A), West-WRF 200-

member ensemble (Figure B-39B), and West-WRF 200-member ensemble + Deep learning (Figure B-39C). The application of the deep learning technique is shown to improve the skill of the raw forecast. It maintains the high precipitation event probabilities while reducing the locational biases. Furthermore, for the Yuba-Feather watershed, the deep learning post-processed West-WRF forecast (shown in purple; Figure B-39D) outperforms all other reference benchmarks, including the ECMWF (green) and the raw West-WRF (red), for the period of assessment (Dec 2021–Mar 2022), from a lead time of 1 to 6 days.



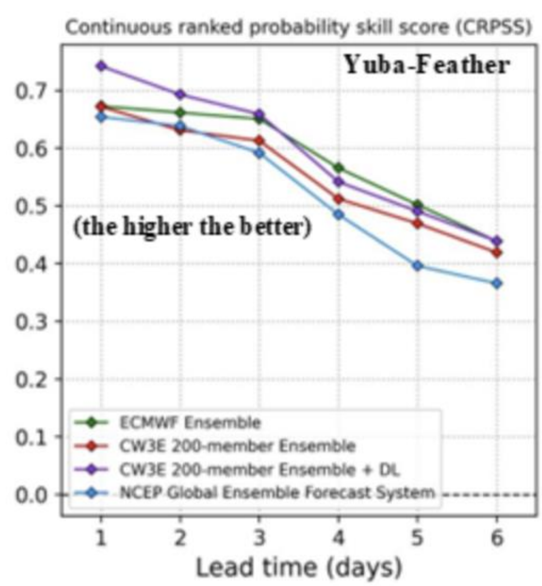


Figure B-39. Skill comparison among GEFS, ECMWF, West-WRF 200-mem ensemble, and West-WRF 200-mem ensemble + deep learning for a representative test case (24 Dec 2021; panels A-C), and for the Yuba-Feather (panel D)

3) Deterministic prediction of IVT in NRT with convolutional neural networks:

This study (Chapman et al. 2019, *GRL*) tests the utility of convolutional neural networks (CNNs) as a postprocessing framework for improving the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS)'s integrated vapor transport forecast (IVT) field in the Eastern Pacific and western United States. Here the forecasts from 3 to 168 hours are examined for the cold season (October–April). GFS forecasts were separated into training (October 2008 to April 2016), validation (October 2016 to April 2017), and testing (October 2017 to April 2018) data sets. IVT from NASA's Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) reanalysis is used as ground truth to diagnose forecast error and for the CNN model training. As shown in Figure B-40, this method reduces errors in terms of RMSE at forecast leads from 3 hours to seven days (5–20% reduction), while increasing the correlation between observations and predictions (0.5–12% increase). This represents an approximately one-to-two-day lead time improvement in RMSE.

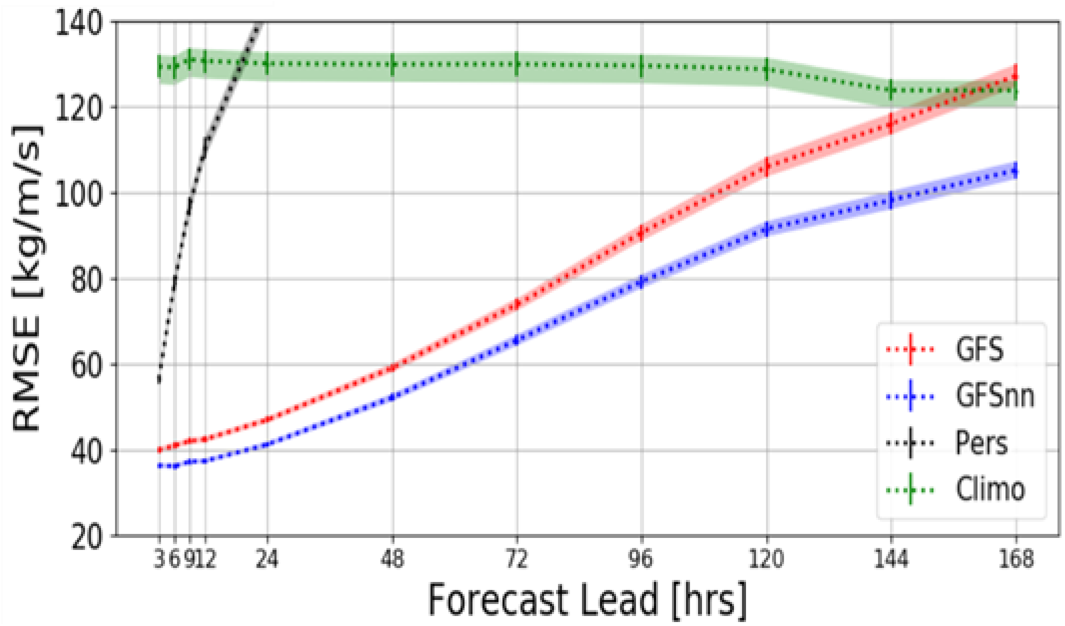


Figure B-40. Temporal evolution RMSE from raw GFS, GFS postprocessed with CNN, Persistence, and climatology forecasts

The CW3E ML team has implemented the associated forecast product for NRT issuance of IVT forecasts for lead times up to 7 days throughout the water year in the Western U.S. (Figure B-41).

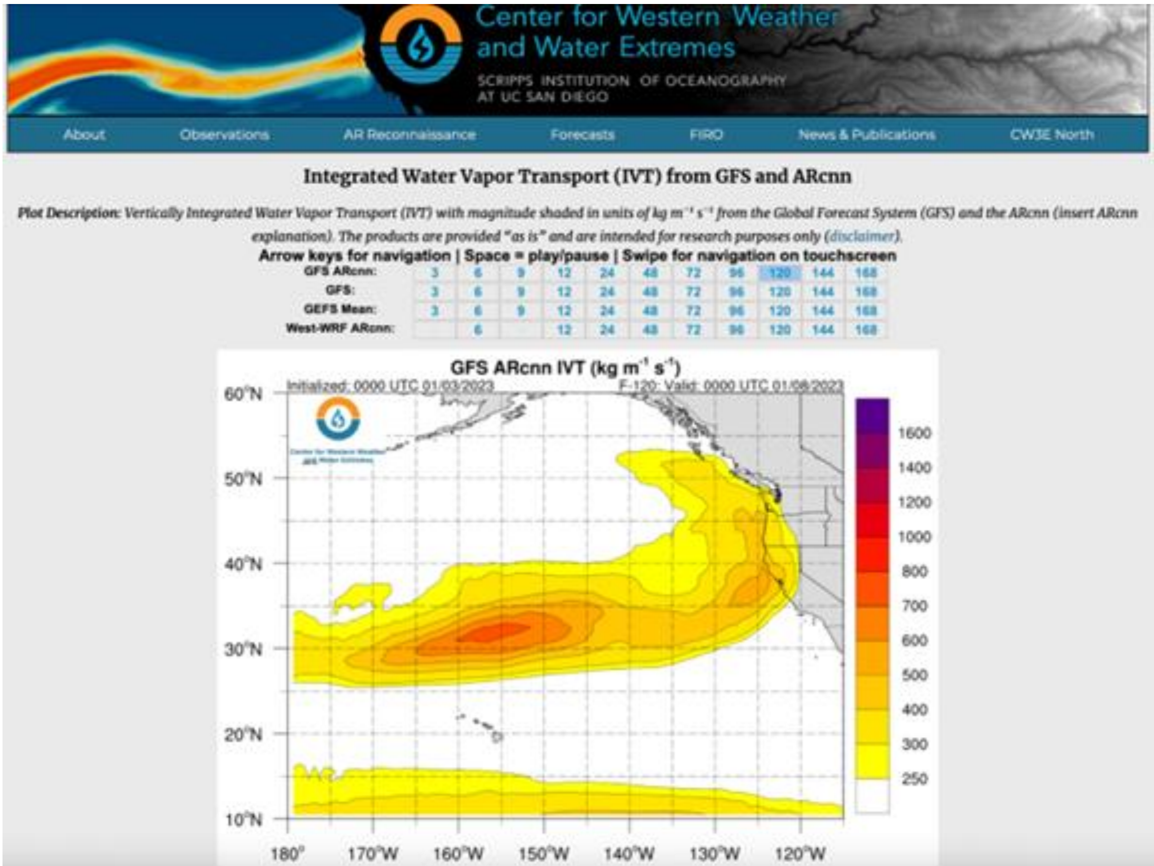


Figure B-41. Forecast product for Deterministic IVT prediction in the Western United States out to 7 days; forecasts are issued throughout the water year via CW3E's near real-time operational system

4) Probabilistic prediction of IVT in NRT with deep learning:

This study (Chapman et al. 2022, *MWR*) explores deep learning postprocessing methods to obtain reliable and accurate probabilistic forecasts from single-member numerical weather predictions of IVT. Using the 34-year CW3E West-WRF Reforecast, dynamically derived 0–120-h probabilistic forecasts for IVT under AR conditions are tested. These predictions are compared with the GEFS dynamic model and the GEFS calibrated with a neural network over coastal locations (Figure B-42; top). In addition, the DL methods are tested against an established, but more rigid, statistical–dynamical ensemble method (the analog ensemble). NASA's MERRA-2 is used as the ground truth dataset. The findings show, using Continuous Ranked Probability Skill score and Brier skill score as verification metrics, that the DL methods compete with or outperform the calibrated GEFS system at lead times from 0 to 48 h and again from 72 to 120 h for AR vapor transport events. In addition, the DL methods generate reliable and skillful probabilistic forecasts.

Figure B-42 (bottom) shows the spread-skill plot of ensemble forecast systems which assesses the ability of the ensemble to quantify uncertainty with the closer to 1:1 line the better. GEFS model (light red) is severely overconfident. The CNN model provides statistically consistent forecasts and indicates that it can capture the flow-dependent forecast uncertainty because its spread dependably reflects the forecast error variance.

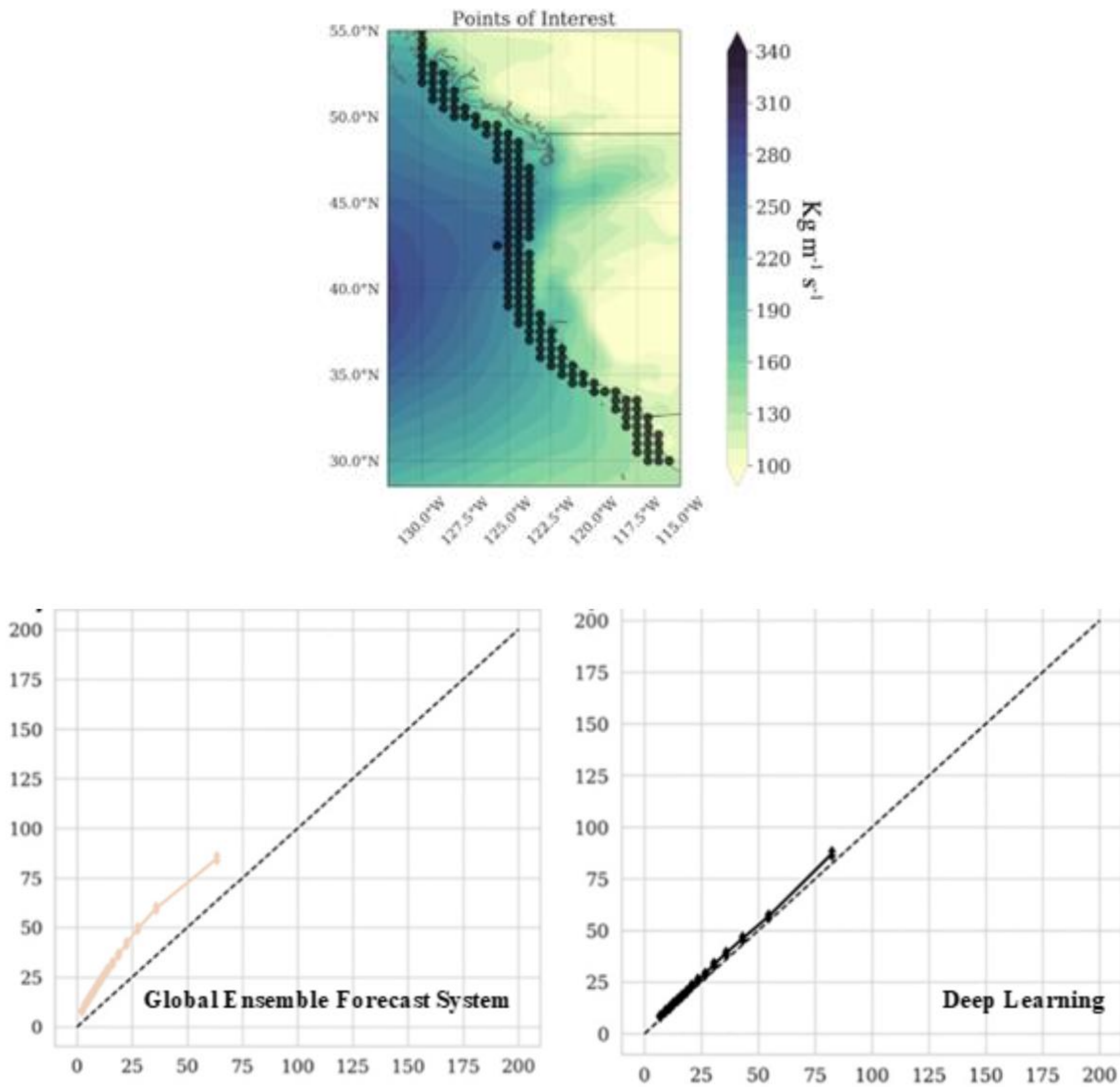


Figure B-42. (*top*) Coastal evaluation locations and climatological (Dec-Mar 1984-2019) IVT (color fill). (*bottom*) Binned spread-skill plots of forecasts from GEFS and CNN model. Horizontal axis represents the binned spread of the ensemble (in kg m1s1) While the vertical axis shows the standard error the mean (in kg m1s1)

B.6. West-WRF and its ensemble: Forecast tool development

Contributing Authors: Brian Kawzenuk, Center for Western Weather and Water Extremes

CW3E has continued to maintain, update, and provide in near-real-time decision support tools to visualize atmospheric processes, including atmospheric rivers, and the resulting impacts on the Yuba and Feather River watersheds. The watershed precipitation forecasting tool has been updated to include additional reservoir catchments, additional models, an optimization on timing for forecast generation, and updated visualizations. The tool now includes Lake Oroville, New Bullards Bar, and Englebright Reservoir catchment areas and forecasts from the GFS, ECMWF, NOAA WPC, National Blend of Models, GEFS, ECMWF EPS, CNRFC, and the West-WRF (two deterministic versions and the 200-member ensemble). Updated visualizations include additional model-to-model comparisons for all the models. Updates to several key ensemble model IVT forecasts were completed to enhance the reliability and timing of these products, as well as expand all models (GEFS, ECMWF EPS, and West-WRF) to three transects along the U.S. West Coast, once which transects the Yuba and Feather River watersheds. Probabilistic and percentile based forecast maps from the West-WRF ensemble are generated to provide additional insight into forecasted quantities of precipitation, snowfall, winds, and temperature over the U.S. West Coast. Point based forecast tools are also generated to give additional details and insight into forecasts from individual ensemble members within the West-WRF.

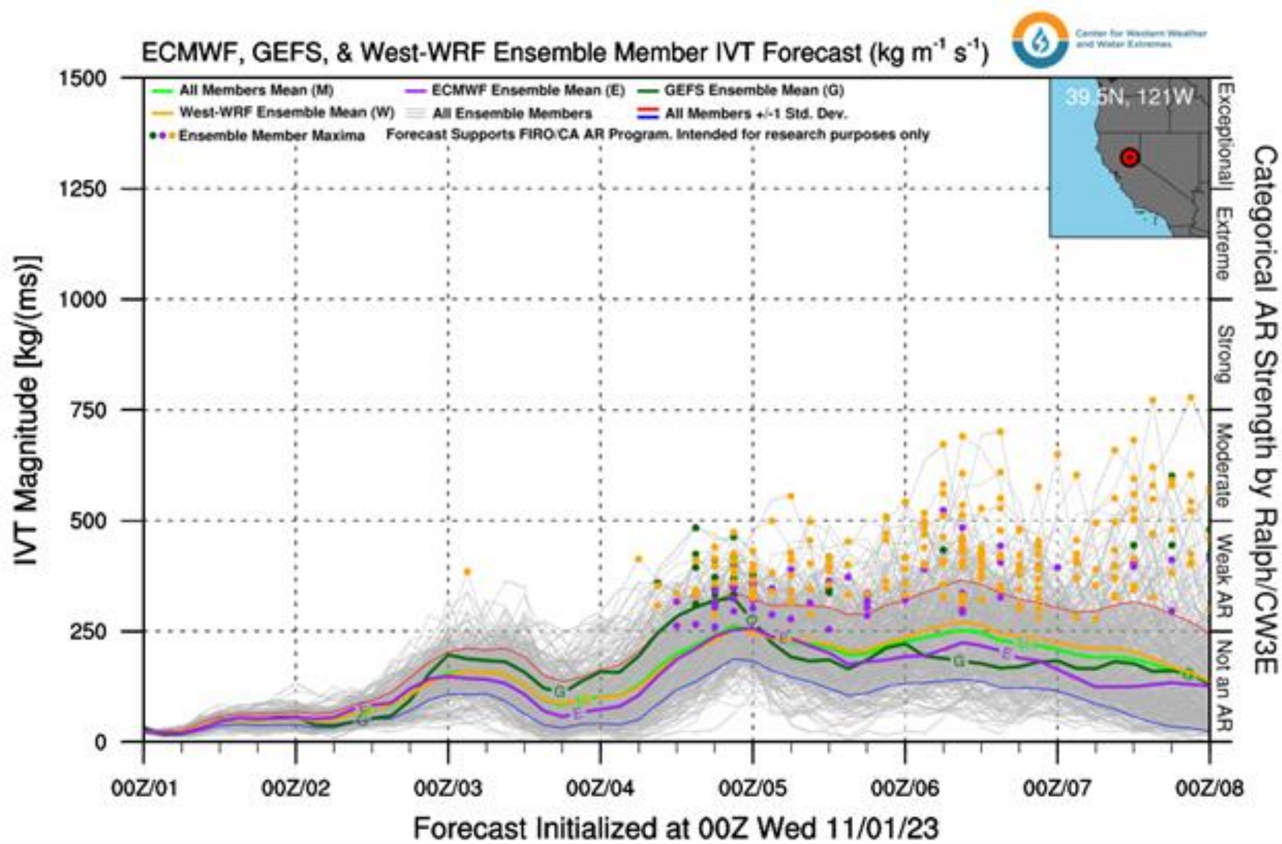


Figure B-43*: Seven day forecast of Integrated Vapor Transport (IVT; kg/m/s) from the GEFS, ECMWF EPS, and West-WRF Ensemble from each ensemble member (thin gray lines), the ensemble means (GEFS; dark green, ECMWF EPS; purple, West-WRF; orange, and all members;*

light green), and plus or minus one standard deviation from the ensemble mean (red and blue lines and gray shading) at 39.5°N 121°W.

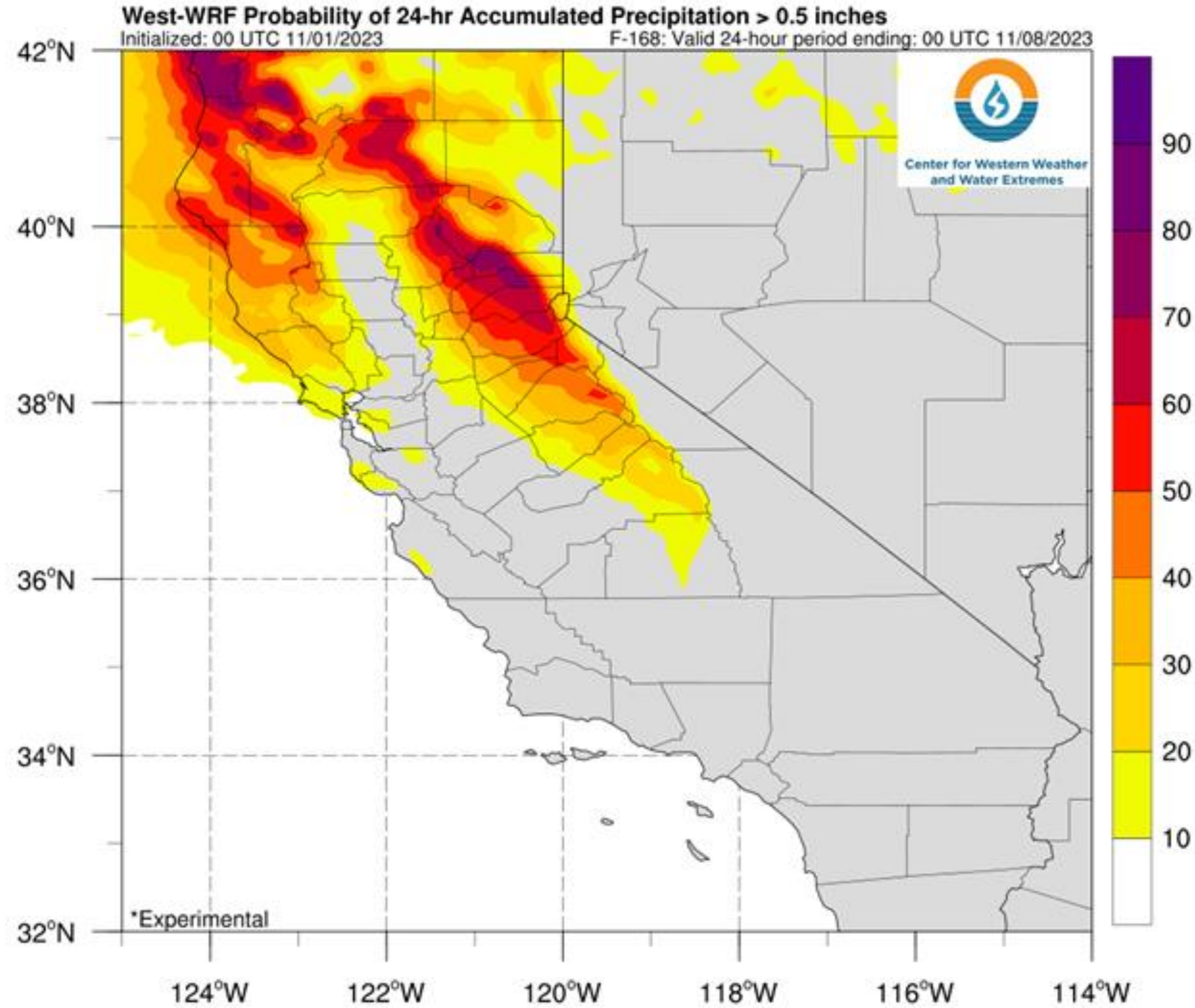


Figure B-44: *Probability of 24-hour precipitation exceeded 0.5 inches from the West-WRF ensemble. Probability is calculated from the number of ensemble members predicting precipitation greater than 0.5 inches at each grid point.*

While extreme precipitation over the Yuba and Feather watersheds is predominantly governed by atmospheric rivers, additional meteorological mechanisms can impact the characteristics, distribution, and forecast performance of precipitation within atmospheric rivers (e.g., narrow cold-frontal rainbands, cutoff lows, etc.). Forecast diagnostics were developed to provide insight into the potential likelihood for such meteorological phenomena that can cause high-intensity precipitation, large-scale precipitation that is not tied to orographic features, and upstream events that may introduce forecast uncertainty downstream.

For example, two-dimensional Peterssen frontogenesis is now included in the suite of high-resolution West-WRF diagnostics to assist in identifying favorable environments for the development of narrow cold-frontal rainbands and, therefore, high-intensity precipitation.

NCFRs generally form and intensify in the vicinity of a strengthening (frontogenetic) cold front and recognizing the potential for these features at longer lead-times will provide essential decision support for an ARs range of impacts.

Additional forecast diagnostics were developed from NCEP Global Forecast System data, including:

- Two-dimensional Pettersen Frontogenesis and Temperature advection for the purposes described above. (Figure B-44)
- Potential vorticity, 250-hPa wind speeds, and irrotational winds to assist in identifying upstream events that can lead to forecast uncertainty over California (Figure B-45)
- 700-hPa Q-Vector and Q-Vector convergence to identify regions where synoptic-scale conditions are favorable for precipitation that is not tied to upslope moisture flux

Figure B-45: *Example forecast image of ECMWF West-WRF two-dimensional Pettersen frontogenesis (shaded; K/100km/3-hr) for the Atmospheric River that impacted Northern California on the 5th of January 2023.*

Figure B-46: *Example forecast image of GFS two-dimensional Pettersen frontogenesis (purple contour; K/100km/3-hr), temperature advection (color shade; K/hr), geopotential height (dam; black contour), and integrated water vapor (mm; gray shade) for the Atmospheric River that impacted Northern California on the 5th of January 2023.*

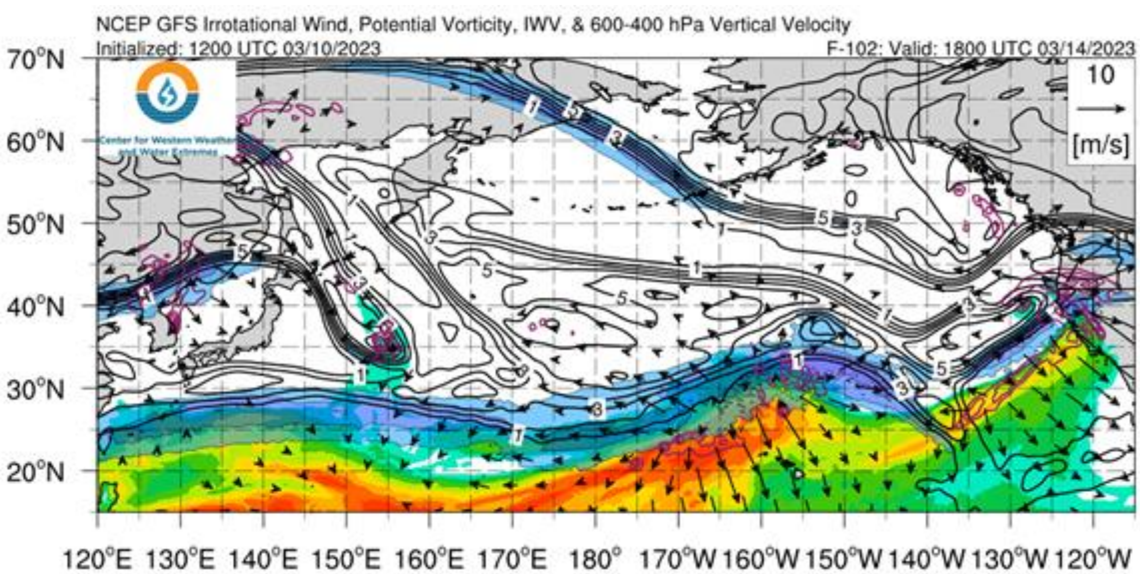


Figure B-47: *Example forecast image of GFS 250-hPa potential vorticity (pvu), 250-hPa wind speed (m/s), 300–200-hPa layer average irrotational wind vectors (m/s), and integrated water vapor (mm) for the atmospheric river that impacted California on the 14th of March 2023.*

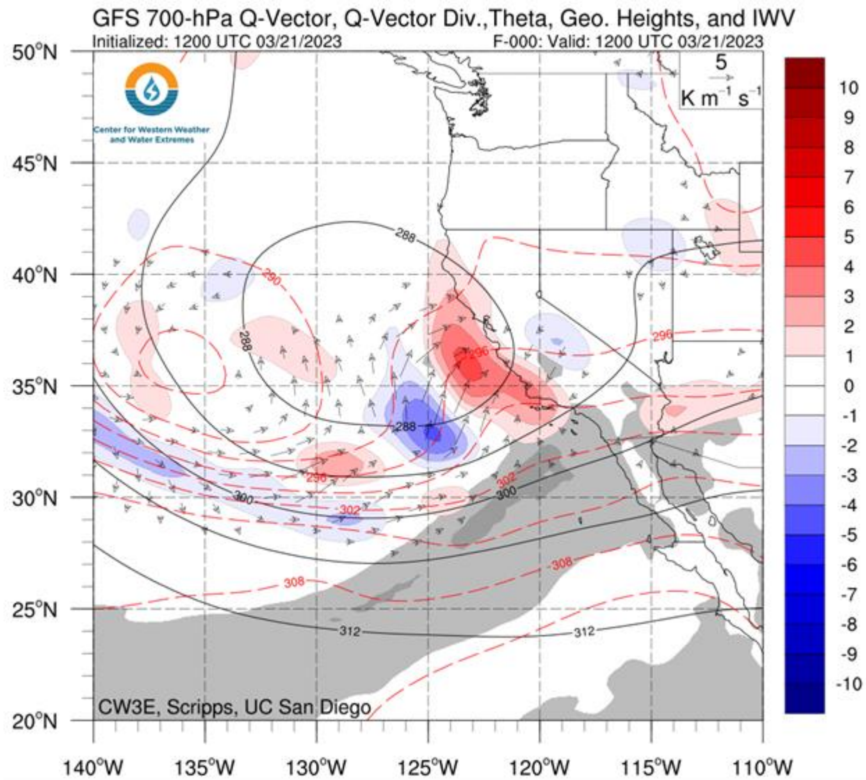


Figure B-48: *Example forecast image of 700-hPa Q-Vectors (vector 10–7 Pa m–1 s–1) , Q-Vector Convergence (10–12 Pa m–2 s–1), potential temperature (K; red contour), geopotential height (dam; black contour), and integrated water vapor (mm, gray shading) for the atmospheric river that impacted Southern California on the 21st March 2023.*

Appendix C – Hydrology

C.1 CNRFC Hydrology Modeling Overview

CNRFC streamflow forecasts are operationally available for the Feather-Yuba system as both five-day deterministic values as well as 365-day ensembles. Both are generated using the NWS Community Hydrologic Prediction System (CHPS) with common model parameters and states. CHPS is an object-oriented modeling framework based on the Deltares Flood Early Warning System (FEWS). The NWS completed its transition from the National Weather Service River Forecast System (NWSRFS) in 2013. CHPS is a combination of integrated and adapted models that describe hydrologic/hydraulic processes as well as a vast array of data and information handling, storage, display, and analysis tools. In the transition to CHPS, common FEWS features were adopted, and a collection of NWSRFS-specific models and tools were adapted into the framework.

The specific models deployed by the CNRFC are quite consistent across their area of responsibility (Figure C-1), but the models for each location are individually configured and calibrated to approximate observed streamflow when presented with observations of precipitation, air temperature, and freezing level elevation. CNRFC models are classified as empirical or "process simulation." They are not rigorous physically based models that attempt to capture the full physics of watershed behavior. CNRFC model applications are "semi-lumped" as opposed to an interconnected grid network. Watersheds with large elevation ranges are typically modeled in two to three elevation bands to better represent elevation-dependent processes, features, and conditions. CNRFC watershed models are run with a six-hour time step and riverine models are run with an hourly time step.

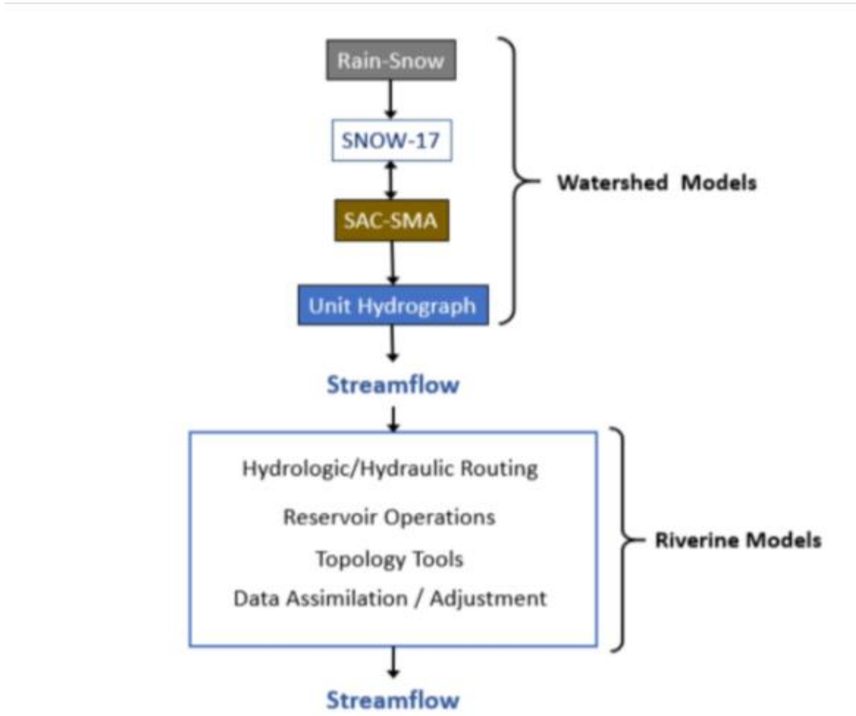


Figure C-1. Watershed and riverine models deployed by the CNRFC within the CHPS framework

The generalized process used to generate the 5-day deterministic forecasts is shown in Figure C-2. Here, the CHPS hydrologic models are presented with new observations and updated meteorological forecasts with each forecast cycle. There is at least one forecast cycle per day (365 days/year), with two on weekdays in the winter and up to four during flood events. As well as needing the latest weather forecast, reliable streamflow forecasts depend on quality control of observations and the monitoring and tuning of model states. In conducting forecasting duties, hydrologists work their way through the model topology for each river basin, making the necessary adjustments to the observational data and model states to achieve (1) a good fit of the simulated streamflow to the observations during the last several days and (2) confidence in the streamflow forecast given the forecast meteorology. When complete, the forecasts are packaged into graphics and text products used to generate public watches and warnings and to help with resource management decisions (e.g., reservoir releases). Current and archived river forecasts can be found on the CNRFC website (www.cnrfc.noaa.gov).

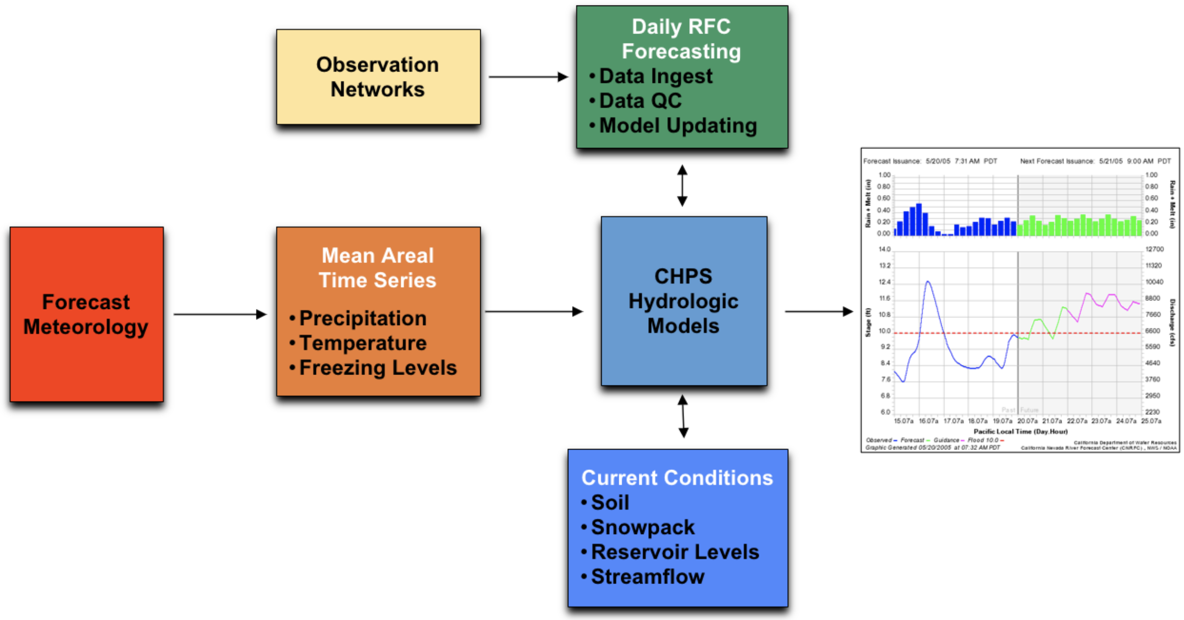


Figure C-2. Generalized forecast process used by the CNRFC to generate five-day deterministic streamflow forecasts

The CNRFC model topology for simulating and forecasting the Feathery-Yuba watershed are shown in Figures C-3 through C-8. This is a highly regulated system, and the CNRFC attempts to model the regulation for the short range deterministic and ensemble (< 30 days) streamflow products. The reservoirs modeled are indicated by squares, and diversions are dashed lines. It should be noted that none of this regulation was accounted for in the hindcast effort but is implemented into the short-range operational streamflow forecast products.

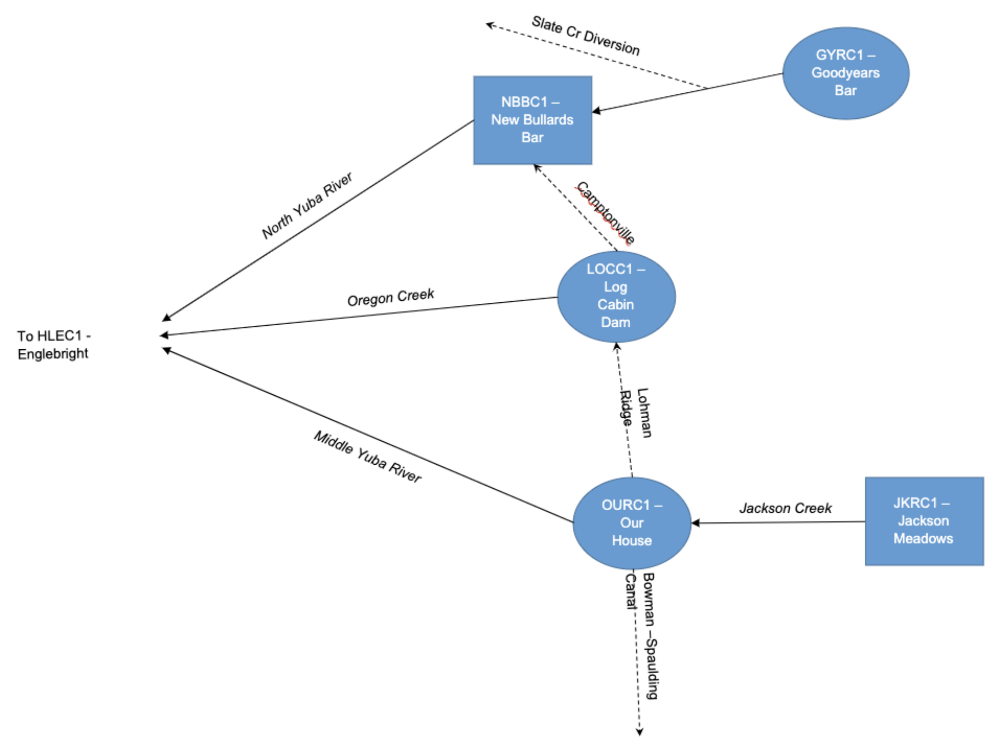


Figure C-3 North and Middle Yuba River Topology

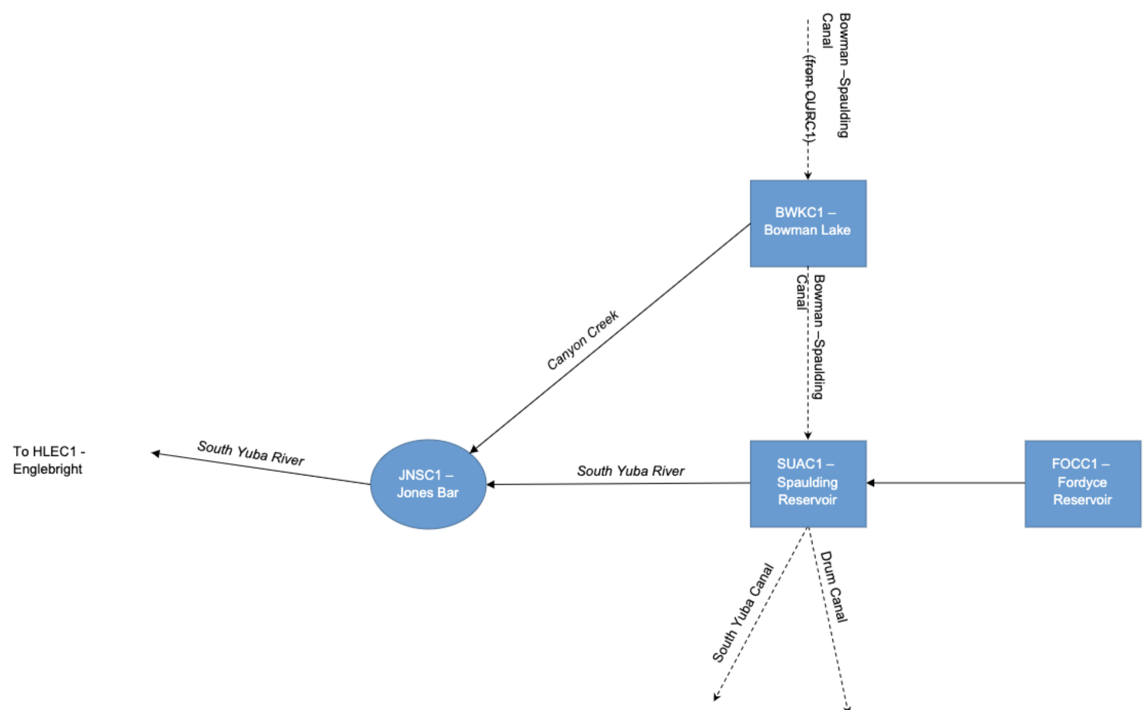


Figure C-4 South Yuba River Topology

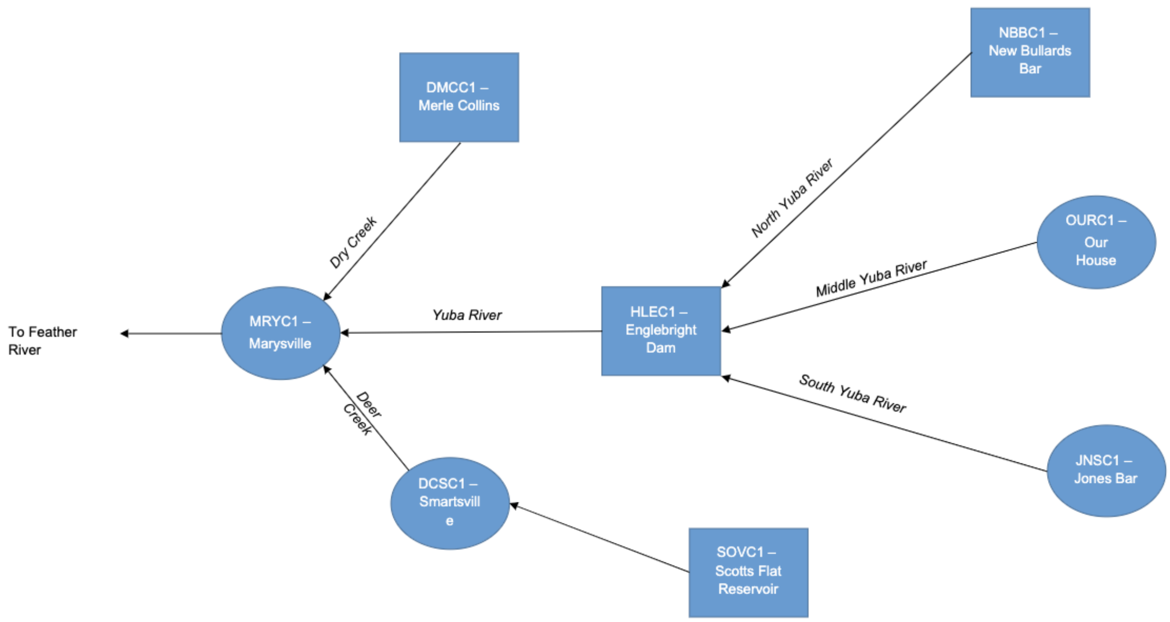


Figure C-5 Lower Yuba River Topology

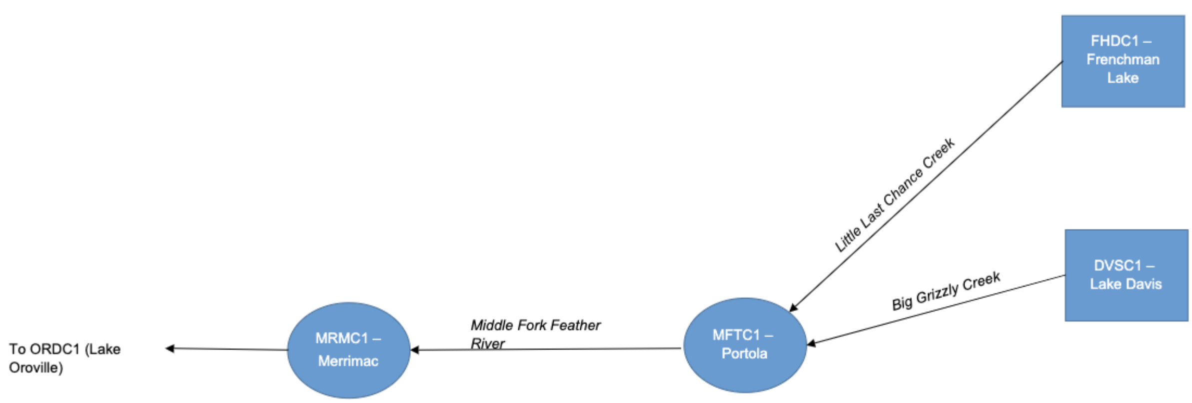


Figure C-6 Middle Fork Feather River Topology

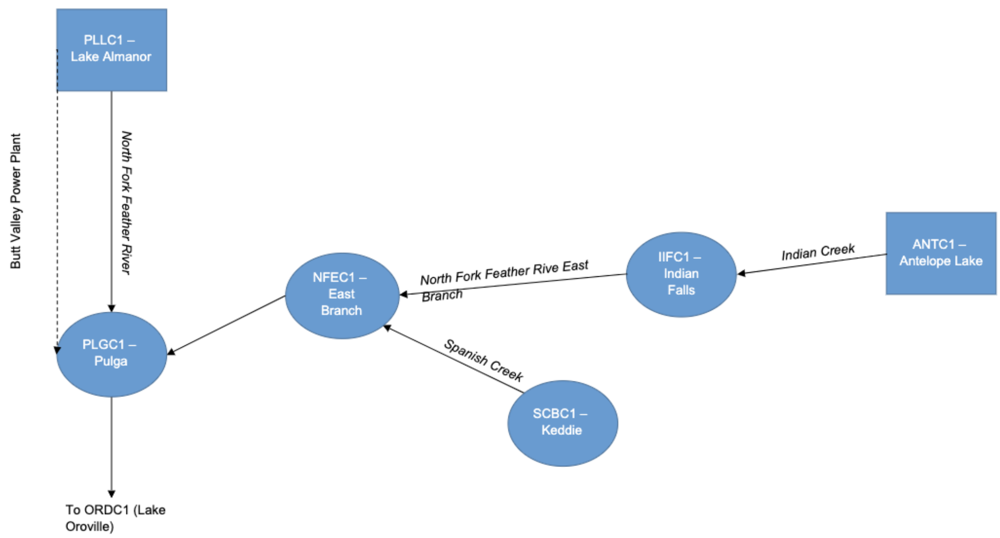


Figure C-7 North Fork Feather River Topology

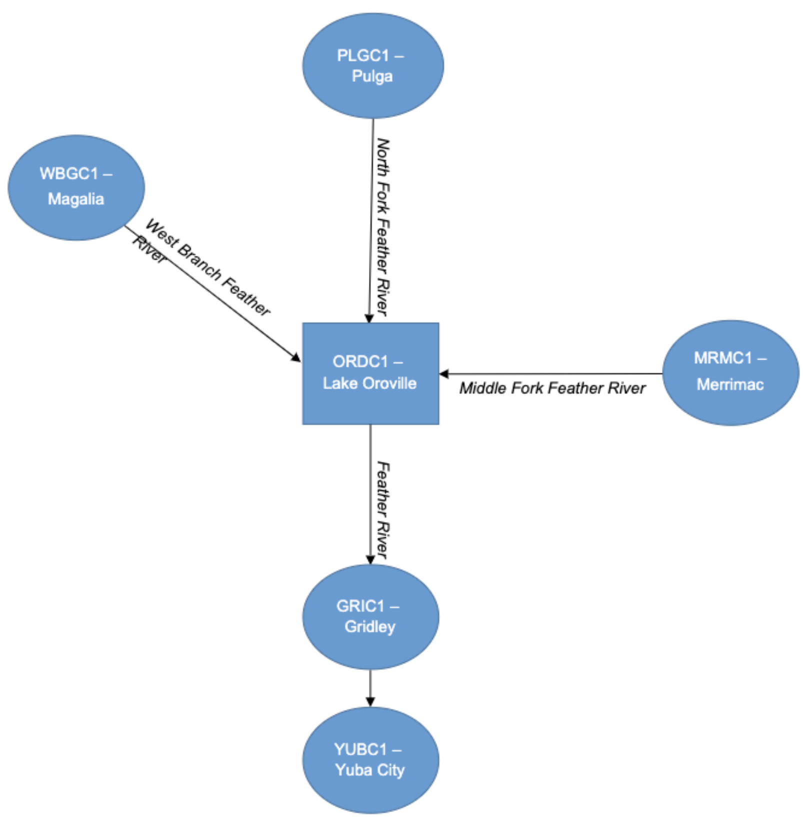


Figure C-8 Lower Feather River Topology

Hindcast Scaling Details

The dates and scale factors associated with the hindcast scaled events are in the following table:

Historical Event	Scale Window	Scale Factor Range
1986	2/15/1986@12:00GMT - 2/20/1986@6:00GMT	1.0 to 1.5 @ 0.1 increments & 1.02, 1.04, 1.06, 1.08, 1.12, 1.14, 1.16, 1.18
1997	12/29/1996@6:00GMT - 1/3/1997@0:00GMT	0.9 to 1.3 @ 0.1 increments & 1.02, 1.04, 1.06, 1.08, 0.84, 0.86, 0.88, 0.92, 0.94, 0.96, 0.98
Mar 1995	3/8/1995@12:00GMT - 3/13/1995@6:00GMT	0.5 through 1.5 @ 0.2 increments
May 1995	427/1995@18:00GMT - 5/2/1995@12:00GMT	0.7 through 1.9 @ 0.2 increments

Table C-1

Appendix D – Observations

D.1 Network evaluation: survey

Metadata for all stations available on CDEC were pulled and sorted for owners and operators of each station for both the Yuba River and Feather River watersheds. Stations were grouped by operator and summarized by the observations collected at each station. Surface meteorology (air temperature, relative humidity), precipitation (rain and snow), and streamflow observations are of particular interest because they are readily integrated into models and forecasts. Streamflow observations were evaluated as part of the Forecast Coordinated Operations program, so this network evaluation and survey focuses on surface meteorology and precipitation. Snow course sites were also left out of this survey since there is a standard protocol separate from in situ stations for how those data are maintained.

Operators maintaining surface meteorological stations and precipitation gauges were surveyed on data quality control, sensors used, whether they operate offline stations, and CW3E website usage. Figure D-1 and Table D-1 summarize the stations and operators assessed and surveyed for the network evaluation.

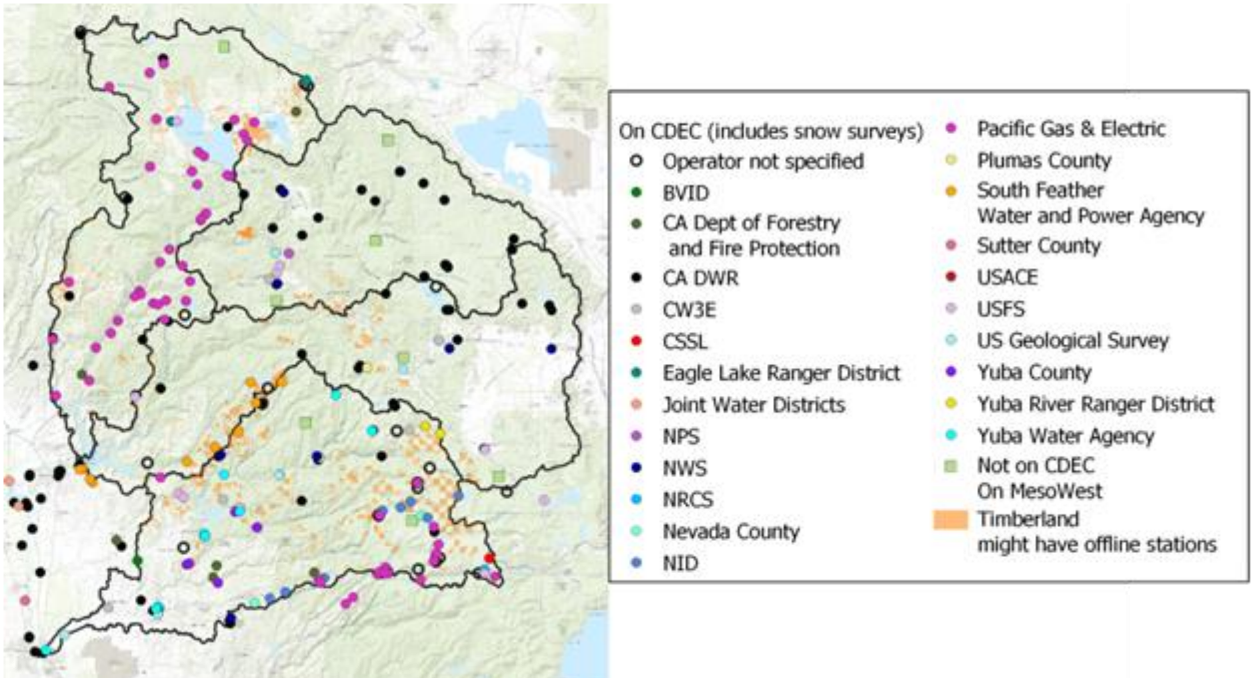


Figure D-1. *All stations included in network evaluation and assessed for network survey colored by operator.*

Operator/agency	Summarized stations available on CDEC?	Survey sent?
Browns Valley Irrigation District	X	X
CA Dept of Water Resources	X	
Central Sierra Snow Lab	X	
Joint Water Districts	X	
National Park Service	X	
National Weather Service	X	X
Nevada County	X	X
Nevada Irrigation District	X	X
Pacific Gas & Electric	X	X
Plumas Corporation	X	
Plumas County	X	
Sierra Pacific Industries	X	X
South Feather Water and Power Agency	X	X

Sutter County	X	
US Army Corps of Engineers	X	X
US Forest Service	X	X
US Geological Survey	X	
Yuba Water Agency	X	X

***Table D-1**. Operators identified for the Yuba River and Feather River watersheds and whether the inventory of stations was summarized and a survey was distributed.*

60% of the operators surveyed responded. This past water year was particularly taxing on operations across the state so many operators were unavailable to respond for a variety of reasons. It is a continued effort to correspond with operators to gather valuable information about their networks.

Respondents reported that stations in their networks experience outages infrequently. Backfilling data from outages and quality control of data are specific to operator and sensor type. 60% of respondents said they QC their precipitation data, however only 20% of respondents send the QCed data to CDEC. The 40% that do not send QCed precipitation data to CDEC reported that there was not a robust process in place for them to send the corrected data.

D.2 Monitoring network updates: CW3E and partners

The USGS, in partnership with DWR, are planning an additional 30 stations to monitor soil moisture in the Feather River watershed. These stations have been scouted using the Basin Characterization Model (Flint et al. 2021) to fill spatial gaps based on landscape characteristics. The USGS soil moisture stations are planned to be deployed over the next couple of years and CW3E is coordinating potentially co-located stations for comparing soil sensors. Our groups are also planning to participate in fieldwork together for knowledge-sharing about each group's respective station protocols.

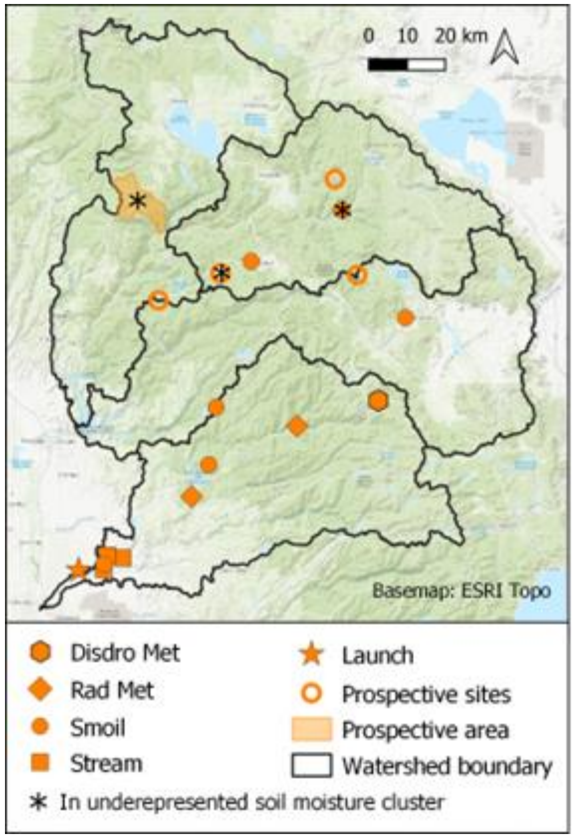


***Figure D-2**. CW3E stations deployed in support of FIRO (solid markers) and prospective stations to continue filling gaps identified through the FIRO process (hollow markers and shaded region). Locations marked with an asterisk indicate sites scouted to fill gaps identified by the cluster analysis in the PVA, particularly the cluster that lacked existing soil moisture stations.*

The remaining SMOIL installations for CW3E include several locations in the Feather River watershed. CW3E is coordinating with USGS and Plumas Corporation to site the stations based on our group's respective permitting strengths, USGS has a multi-forest agreement with USFS and CW3E has the benefit of being able to permit with private landowners. The locations marked with asterisks in Fig. D-2 highlights the areas where CW3E can or has pursued station locations with private landowners to help fill spatial gaps in soil moisture monitoring in the Feather River watershed.

Name	Watershed	Code	Latitude	Longitude	Elevation (m)	Station Type	Installation date
Skyline Harvest[S]	Yuba	SKY	39.470969	-121.091673	833	SMOIL	Oct 2019
Northstar Meadow*	Yuba	NSM	39.605249	-121.071594	1235	SMOIL	Aug 2020
Lower Bathhouse (SFSU)[H]	Yuba	LBH	39.624073	-120.577654	1680	Disdro Met	Oct 2020

Downieville[H]	Yuba	DLA	39.5634	-120.8242	901	Rad Met	Oct 2019
New Bullards Bar Dam[H]	Yuba	NBB	39.396359	-121.1437698	634	Rad Met	Dec 2019
Feather River College	Feather	FRC	39.945873	-120.969701	1044	SMOIL	Nov 2019
Marysville, Kibbe Road	Yuba	USYUB	39.220808	121.482356	30	Launch	2019
Truckee radar	Martis Cr.	TRK	39.328435	120.122274	1789	MRR	Mar - June 2020
Browns Valley School	Yuba	BVS	39.23586	-121.40621	71	SMOIL	Apr 2021
Portola	Feather	POR	39.8175	-120.4969	1509	SMOIL	Oct 2021
Sycamore Ranch*	Yuba	SYR	39.22389	-121.407016	46	Stream	Aug 2021
Little Dry Cr.*	Yuba	LDM	39.256644	-121.39706	65	Stream	Aug 2021
Upper Dry Cr.*	**Yuba**	**UDC**	**39.251277**	**-121.350107**	**236**	**Stream**	**Aug 2022**
Heart K Ranch	**Feather**	**HKR**	**40.067822**	**-120.693752**	**1150**	**SMOIL**	**Aug 2023**

Table D-2. *Observations added in support of FIRO objectives. * = not telemetered (as of September 2023). [S] = non-standard soil pit depths because bedrock was reached. [H] = Heated tipping bucket. New stations installed post-PVA are in bold.*

References:

Flint, L.E., Flint, A.L., and Stern, M.A., 2021, The basin characterization model—A regional water balance software package: U.S. Geological Survey Techniques and Methods 6–H1, 85 p., https://doi.org/10.3133/tm6H1.

D.3 Radiosonde sampling

CW3E has been conducting winter radiosonde sampling from Marysville, CA (USYUB) since water year 2021. The USYUB radiosonde launch location is located near the foothills of the Sierra Nevada just west of the Yuba River watershed. Radiosondes collect vertical transects of air temperature, humidity, air pressure, and wind speed and direction during their ascent and descent and the data are sent directly to the Global Telecommunications System upon completion of each launch. These data are then readily assimilated into weather forecast models worldwide. Since installation, 208 radiosondes have been launched from USYUB (see Figure D-2 for water year 2023 radiosonde trajectories). Radiosonde sampling is also a key component of AR Reconnaissance, and the observations are used in efforts to improve models and forecasts (see section 5.2.3.c AR Reconnaissance for more detail).

A subset of radiosonde launches in water year 2023 were called to investigate how representative the WWRF freezing level forecasts within the Yuba watershed were to the radiosonde freezing level. This is ongoing work to be completed post-FVA.

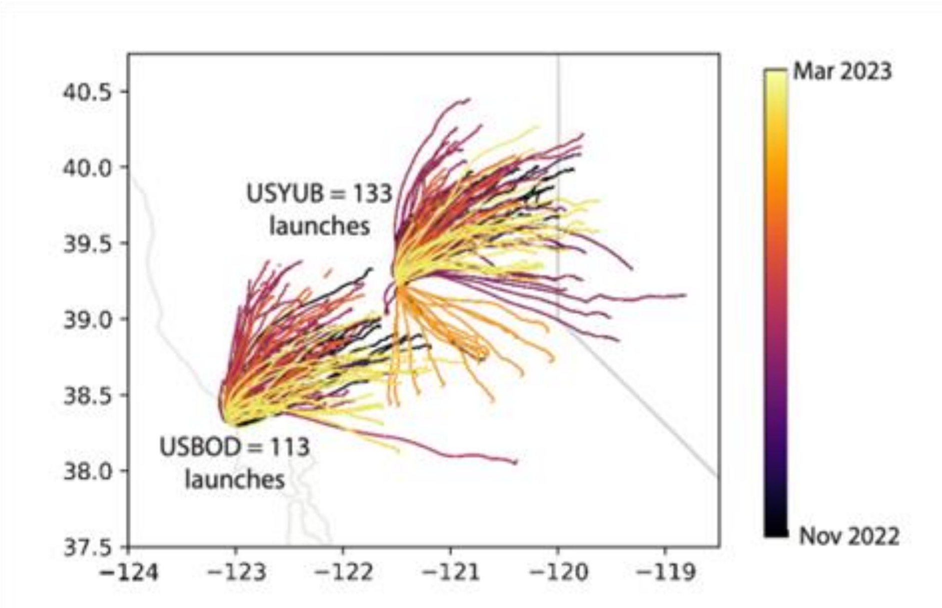


Figure D-2. *Radiosonde trajectories for all radiosondes launched during the water year 2023 winter sampling period. Flight paths colored by date (black = earlier in the season, yellow = end of the season). USYUB is the Marysville, CA launch location. 133 radiosondes were launched for the 2023 season.*

D.4 High elevation precipitation

Both the Yuba and Feather watersheds have an extensive snow observational capacity. The Feather watershed has 10 snow pillows that measure the snow water equivalent (SWE) on an hourly basis. Many of these stations are also being upgraded to sample the snow on a 15-minute basis. Snow pillows across the Feather sample a large elevational gradient with a maximum elevation of 8,338 ft (Lower Lassen Peak) and a minimum of 5,202 ft (Four Trees). The average April 1st SWE at these locations are 73” and 19.3”, respectively. Note that while Four Trees experiences less snow accumulation, it is not necessarily drier but rather has a higher rain-to-snow ratio than Lower Lassen Peak.

Measurement errors at many of the telemetered snow pillows necessitate occasional winter “control” measurements. Control measurements are taken using a Federal Sampler, usually at all four corners of the snow pillow. The measurements are averaged and the difference between the snow pillow SWE and the control is sometimes used to adjust the reported snow pillow observation. Of the 10 snow pillows, only Lower Lassen Peak and Pilot Peak have not received control measurements over the past 5 years. Note that this may be due to their remote locations.

Of the 10 snow pillow sites, 4 locations also have snow courses. Snow courses predate snow pillows by decades and involve a snow surveyor sampling the snow with a Federal Sampler—usually once a month starting in January and ending in April—across a transect of approximately 10 points. The snow courses at snow pillow sites can be used as a "control" for snow pillow sites. In total there are 25 snow courses in the Feather basin. Online snow course records begin in 1930, with the lowest sitting at 4,600 ft (Chester Flat), and the highest at 7,449 ft (Mount Dyer 1). The mean April 1st SWE across all 25 snow courses is 23.16".

The Yuba has 4 snow pillow sites, with the lowest-elevation station (Sunnyside Meadow, 6,300 ft) installed most recently in October 2019. The highest station sits at 7,200 ft (Meadow Lake), with an average April 1st SWE of 48.7". In general, Yuba snow pillow sites are "higher" than in the Feather (by about 100 ft), but the average April 1st SWE is roughly 23% greater. However, the Yuba snow pillows sample a narrower fraction of the watershed hypsometry than the Feather pillows. Roughly 80% of the Yuba basin area exists at elevations below the lowest snow pillow (compared to 50% in the Feather), and 5% lies above the highest snow pillow (compared to 1% in the Feather).

Control measurements are rarely taken in the Yuba. The Central Sierra Snow Laboratory (a UC Berkeley facility) snow pillow and snow survey are owned and operated by the Natural Resources Conservation Service (NRCS) as part of the Snow Telemetry (SNOTEL) network and are therefore not subject to the DWR control measurement program. Only Meadow Lake experiences regular control measurements of the four stations. However, while control measurements are sparse, the Yuba does have a relatively more expansive snow course network—particularly in the southeastern quadrant of the basin. Snow courses occur as low as 4,850 ft and extend to 7,800 ft covering a larger extent of the basin hypsometry than the snow pillows. The mean April 1st snow course SWE is 33.7", with the earliest measurement dating to 1910 and an average start date of 1939. Note that while snow pillow observations benefit from frequent sampling to assess storm-scale changes to snowpacks, the less-frequently sampled snow course observations benefit from spatially broader and consistent measurements that can more comprehensively assess the snowpack storage in a given water year to support spring-summer streamflow forecasting.

Metadata and geolocation of snow pillow stations remain problematic in both the Feather and Yuba watersheds (Table D-3). For example, most stations could not be reliably "geo-identified" using high-resolution satellite imagery (Google Earth, NAIP, etc.). Geolocation errors naturally contaminate comparisons of stations to gridded datasets. Pictures of snow study sites (like NRCS) would not only aid geo-identification but also help to improve our understanding of the station "layouts" and representativeness. For example, collocated snow depth and SWE measurements are critical to proper bulk snow density estimates. Moreover, whether snow pillow sites are densely surrounded by trees (or not) deeply affects the energy balance of the snow, which is fundamental to model comparison/evaluation. Finally, knowledge of instrument changes (or upgrades) is also key to understanding a site's ability to measure atmospheric river activity. For example, shifting from acoustic snow depth sensors to laser-based snow depth sensors is critical to capturing storm timing and evolution. While researchers can subjectively "guess" when instrument changes have occurred due to changing data quality, having the metadata to confirm these switches is crucial.

Basin	CDEC code	Latitude	Longitude	Location Google Earth verified?	Elev. (m)	Sensor notes
Feather	LLP	40.466602	-121.50811	**Can't verify**	2541	Temp (hourly)), RH (hourly), Sdepth (event)
Feather	KTL	40.14	-120.715	**Can't verify**	2225	Precip (hourly), Temp (hourly)
Feather	GRZ	39.917	-120.645	**Can't verify**	2103	Precip (hourly), Temp (hourly)
Feather	PLP	39.785892	-120.877777	Accurate	2073	Precip (hourly), Temp (event), Sdepth (event)
Feather	GOL	39.674767	-120.617167	Accurate	2057	Precip (hourly), Temp (hourly)
Feather	HMB	40.115	-121.368	**Can't verify**	1981	Precip (hourly)
Feather	RTL	40.12791	-121.04397	Accurate	1893	Precip (hourly), Temp (hourly)
Feather	HRK	40.418	-121.275	**Maybe**	1890	temp (hourly), Sdepth (hourly), RH (hourly)
Feather	BKL	39.85292	-121.25135	**Maybe**	1790	Precip (hourly)
Feather	FOR	39.81278	-121.32168	**Maybe**	1586	Precip (hourly), Sdepth (event|2022), Temp (event)
Yuba	CSL	39.325	-120.367	Accurate	2103	Precip (hourly), Temp (hourly), Sdepth (hourly @nrcs), Soil moist (hourly @nrcs)

Yuba	MDW	39.405663	-120.506058	**Maybe**	2195	Temp (event), RH (event), Sdepth (event)
Yuba	RCC	39.621864	-120.679871	**Maybe**	1975	Temp (event), RH (event), Sdepth (event)
Yuba	SSM	39.698132	-120.782211	**Maybe**	1920	Temp (event), Sdepth (event), RH (event), Soil moisture (event)

Table D-3. *Snow pillow metadata for stations in the Yuba River and Feather River watersheds.*

Finally, continuing to build out snow pillow observation sites with disdrometers, laser-based snow depth, and soil moisture sensors is critical to understanding of the role of atmospheric rivers on snow and downstream hydrologic processes. The presence of disdrometers removes the need for precipitation phase proxy measurements; laser-based snow depth provides accurate and precise measurements of snowfall accumulation timing and magnitude; and soil moisture aids in the understanding of whether snow is "absorbing" or "transmitting" liquid water to the land surface during intense, warm atmospheric river events.

D.5 Freezing level

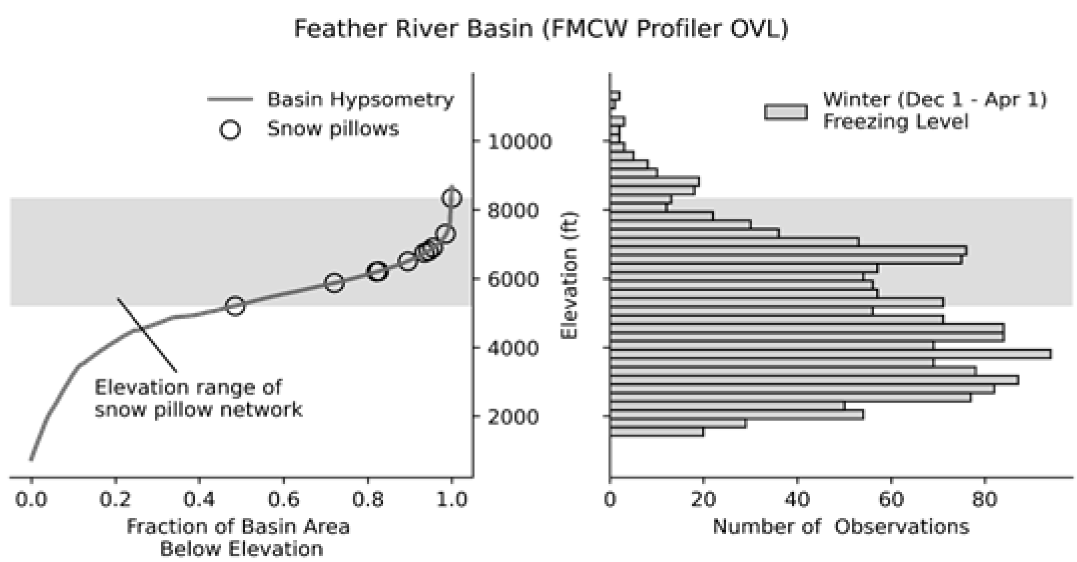


Figure D-3*. Hypsometry of the Feather River watershed with locations of snow pillows (circles) plotted (left) and histogram of observed freezing levels from NOAA's FMCW radar in Oroville, CA (374 ft elevation) over the last five winters binned by elevation (right).*

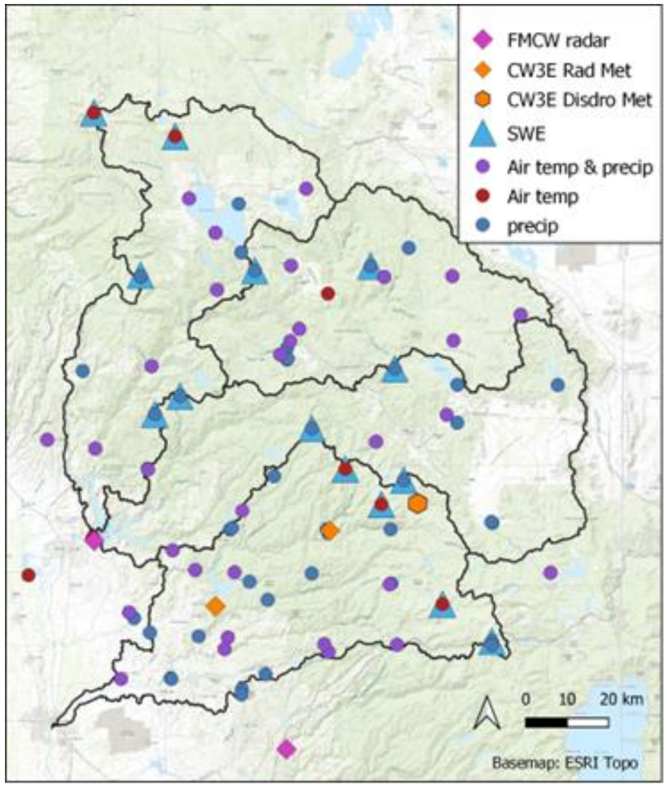


Figure D-4. *Stations referenced for freezing level verification. FMCW radars (magenta diamond) and CW3E MRRs (orange diamond) measure snow level in the atmosphere. CW3E disdrometers (at orange diamonds and orange hexagon) measure precipitation phase at the surface. Additional surface observations referenced include SWE (blue triangles), air temperature (red and purple dots), and precipitation (blue and purple dots). SWE stations with collocated instrumentation have the markers overlaid.*

D.6 Quantitative precipitation estimation

We obtained the raw hourly precipitation data from CNRFC and perform a quality control (QC) for all the hourly gauges following CNRFC QC procedures with minor adaptations. The raw precipitation gauge data sources include: the Geostationary Operational Environmental Satellites (GOES) Data Collection Platforms (DCPs) [data received from the NOAA Hydrometeorological Automated Data System (HADS)], ALERT (Automated Local Evaluation in Real Time) system (data collected locally at the county/water agency), the California Department of Water Resources (CA DWR)/the California Data Exchange Center (CDEC) gauges, SNOTEL data (queried from the NRCS database by MADIS), as well as the Automated Surface/Weather Observing System (ASOS/AWOS) gauges. The monthly precipitation from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994, 2008) is used to define the background climatology.

As part of the QC process, we calculated ratios between hourly gauge precipitation and the PRISM climatology at station grid nodes, following the MM method. For a target gauge, we use nearby 5 stations with non-missing and non-bad observations within 100 km. We interpolate their ratios onto the target station grid using the IDW method. If the normalized deviation of the target gauge value (from the estimation by nearby gauges) is within a certain threshold, which is selected by a sensitivity test, the observation is considered passing the QC.

Similarly, the gridded hourly precipitation estimation can be obtained by multiplication of the gridded ratios with the PRISM climatology. CNRFC adopts manual screening for the 6-hourly QC'd precipitation product after the automatic QC procedures.

The tool uses precipitation gauges that are used in the CNRFC's 6-hourly QPE product to ensure that our products are comparable (see Figure D1 for gauge locations; raw gauge data were sourced from multiple databases (See Appendix D)). The Mountain Mapper (MM) method, uses an inverse distance weighting (IDW) approach to estimate precipitation at a certain location within the domain by considering the climatology of precipitation and observations from gauged locations. Hourly precipitation data were quality controlled and evaluated against the precipitation climatology (see Appendix D). The 6-hourly gauge data produced by CW3E were compared to CNRFC's 6-hourly gauge data to ensure the performance looks reasonable (see Figure D2).The QCed gauge precipitation data are an input to the MM-based QPE product.

The performance of the Machine Learning (ML) product was assessed via root-mean-square error, spatial correlation, and bias in which it outperformed both IDW and MM.

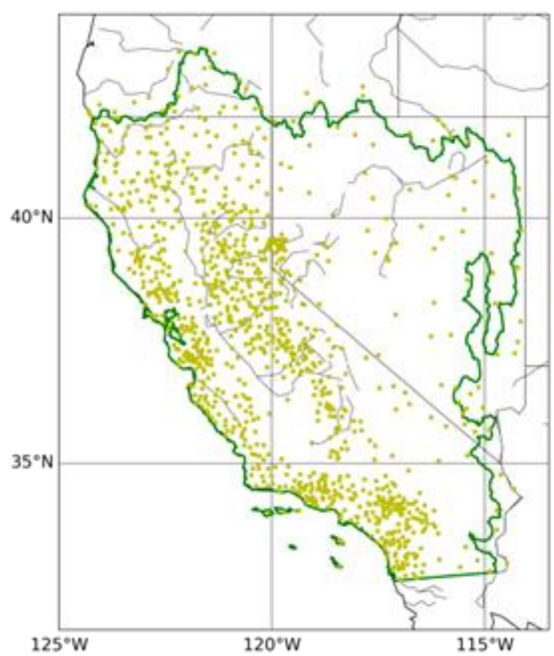


Figure D-5. *Map of hourly precipitation gauges*

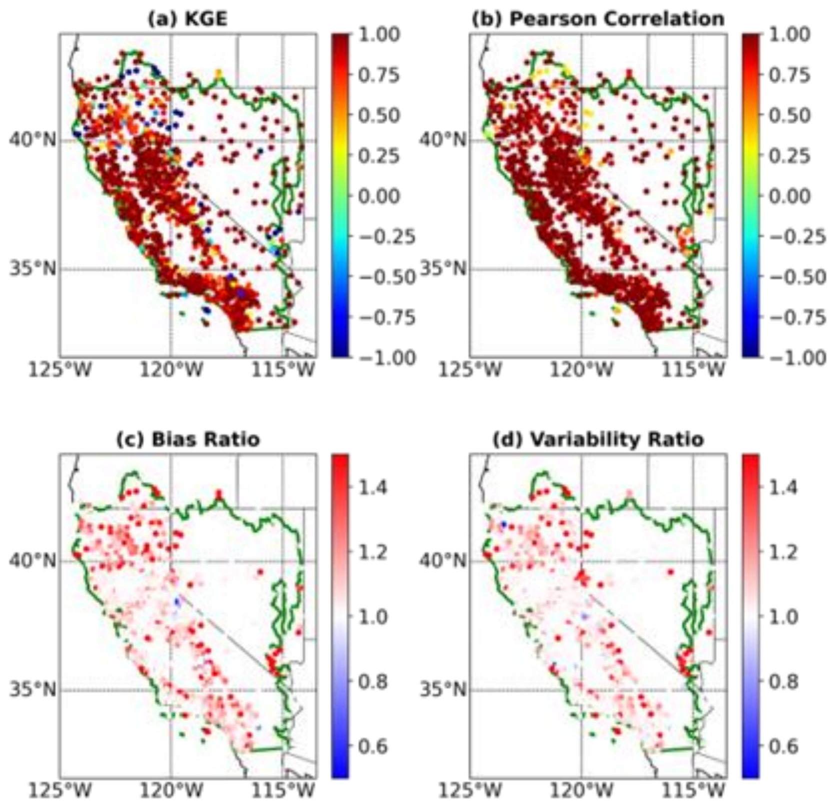


Figure D-6. *Hourly quality-controlled precipitation aggregated to 6-hourly versus CNRFC 6-hourly quality-controlled precipitation*

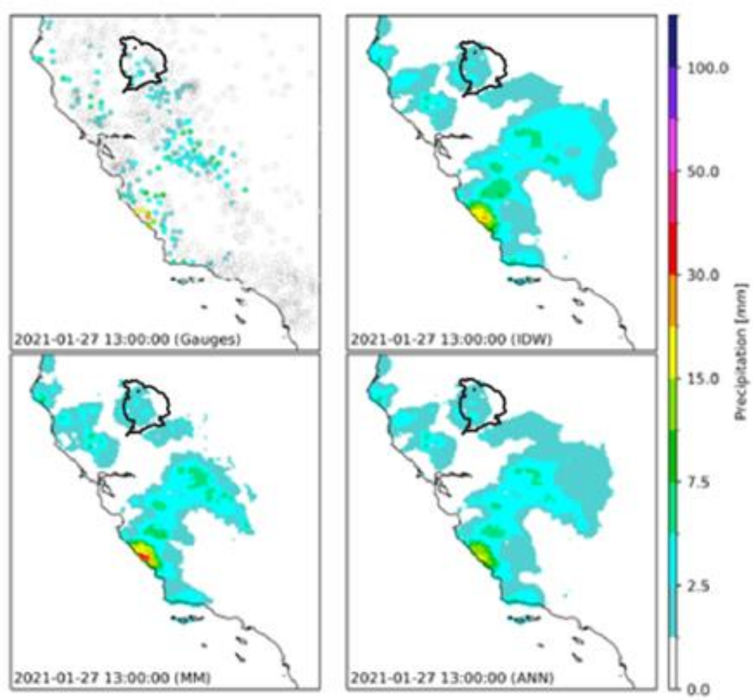


Figure D-7. *An example of the gridded QPE developed by different methods on 2021/01/27 at 1 PM. The four panels show quality-controlled gauge data, QPE products developed by Inverse Distance Weighting (IDW) interpolation, Mountain Mapper (MM), and Machine Learning (Artificial Neural Network (ANN)), respectively. Yuba River and Feather River watersheds are outlined.*

Appendix E – Verification

E.1. QPF error tendencies

Deterministic forecast evaluations of global and regional forecasts were provided in the Yuba-Feather PVA. In short, top 10 percentile events using 72-hour precipitation were shown to have skill metrics exceeding appropriate thresholds out to 6–8-day lead time. The West-WRF reforecast (the regional model), had better variance explained but had larger random errors, suggesting that some bias correction could be made to improve forecasts. Many of the statistics imply the overall skill of the weather modeling system but could be skewed based on the influence of a few key events. This section describes the frequency of forecast errors and how overforecasts and underforecasts contribute to the overall error patterns.

This investigation utilizes the CW3E West-WRF reforecast (Cobb et al. 2022), which contains 34 years of precipitation forecast data at a 3 km resolution over California, Oregon, and southern Washington. Mean areal precipitation (MAP) is computed over the Upper Yuba River and Feather River watershed at the HUC8 level between 1 December and the following 31 March for lead times of 1 through 5 days. The forecasts are compared to observations from the Stage-IV quantitative precipitation estimate (QPE) from 2004 to 2018. Forecast errors are computed on 24-hour totals and the tendencies of total seasonal error, occurrences of overestimations and underestimations, and the percentage of error from the top 3 bust forecasts to the total forecast error. Similar results are found for the North Fork of the Feather (figures E-3 and E4).

In summary, the QPF errors on average tend to be consistently more overestimated than underestimated over the Upper Yuba River basin across all lead times (Figure 5-14) and the impact of the overestimated QPF is larger than the impact on underestimating QPF (Figure 5-15) with respect to total QPF error. This result is most robust at the 2 and 3-day lead times of total QPF error (where the black vertical lines do not overlap between the overestimations and underestimations) in the number of events and for 1 and 2-day lead times for total QPF error. At lead times where they do overlap, the data suggest that there are some water years where the overestimations and underestimations are more similar in frequency and impact. It is important to understand these error tendencies because it is a pivotal first step in the process to improve model forecasts. Trends in the model forecasts can be isolated, identified, and investigated for future model skill improvements.

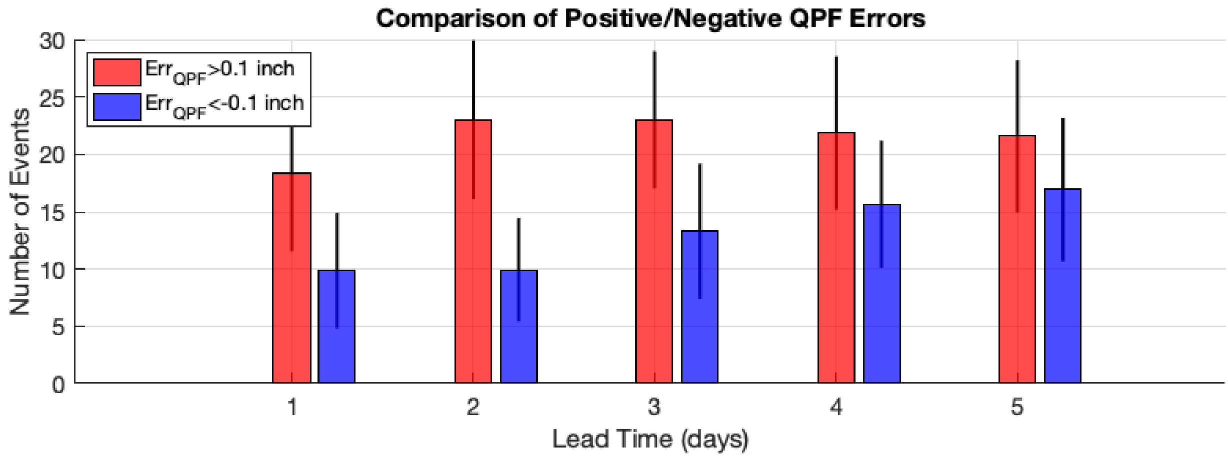


Figure E-1. *Frequency of overestimated (red) and underestimated (blue) MAP QPF error over the Upper Yuba basin as a function of forecast lead time (days). The height of the bars represents the average error of the winter season (Dec-Mar) over a 14-year period, and the black vertical lines represent one standard deviation around the mean.*

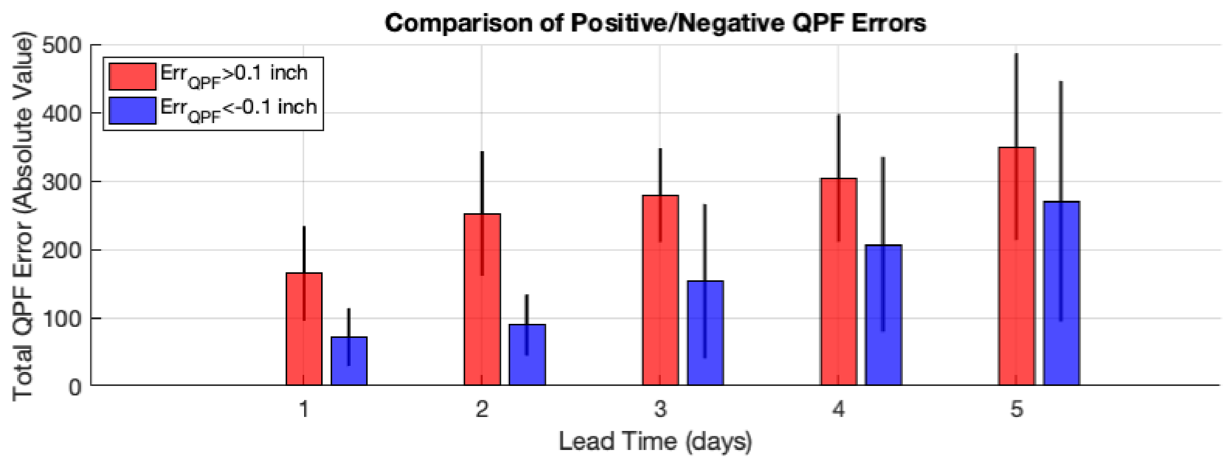


Figure E-2. *Same as Figure E-1 for MAP QPE total error over the winter season from Dec-Mar.*

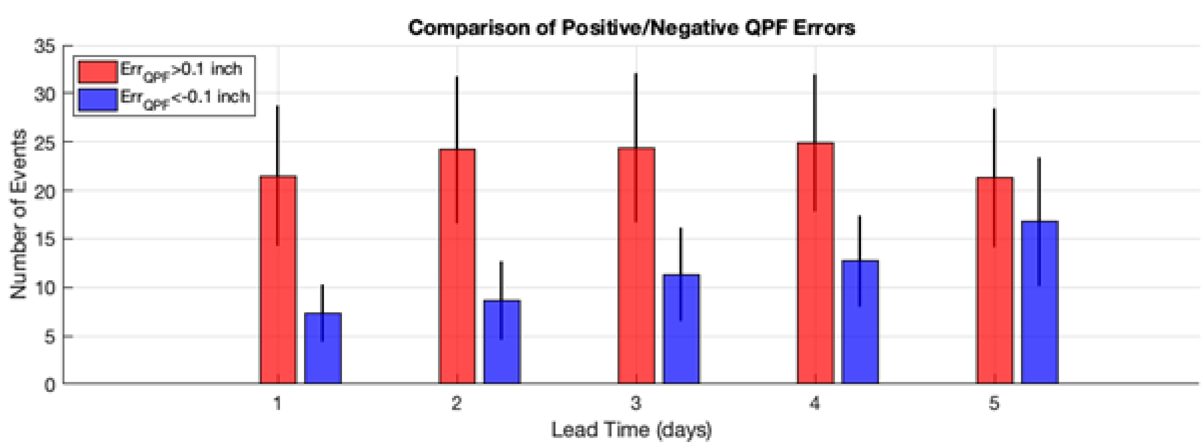


Figure E-3. *Same as Figure E-1 for the North Fork Feather watershed.*

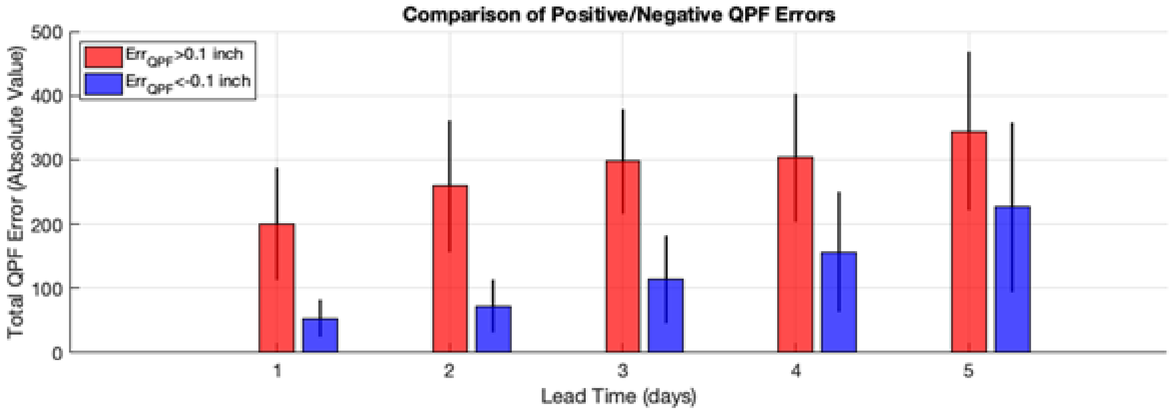


Figure E-4. *Same as Figure E-2 for the North Fork Feather watershed.*

E.2. Freezing Level Evaluation

The partitioning of rain and snow remains an important forecasting challenge for areas of the Sierra Mountains due to the hydrologic impacts on runoff generation and snowpack accumulation during precipitation events. Forecasting precipitation as frozen vs. liquid has significant implications for water management strategies. Freezing levels (ZFL) are often used as a proxy for indicating where frozen precipitation might occur as it explicitly represents the altitude of the 0°C isotherm of the vertical temperature profile. As reported in the PVA, Sumargo et al. (2020) found inflow volume uncertainties of under 10 percent to over 50 percent of the flood pool storages at the Oroville and New Bullards Bar, depending on the ZFL, antecedent moisture condition, and precipitation event magnitude, using an average ±350 m ZFL forecast error.

Forecasts of ZFL at Oroville and Colfax were previously assessed during the Yuba/Feather PVA using archived near-real time CNRFC data and Frequency-Modulated Continuous Wave (FMCW) vertically profiling radars (Johnston et al. 2017). However, this assessment leveraged FMCW brightband heights, or altitude of the maximum radar reflectivity, as observations and thus only represented forecast skill of above-terrain freezing levels. It was recommended that forecasts correctly predicting freezing conditions at the surface should also be examined to convey the skill of snowing conditions.

The discrepancy between the characteristics used to define partitioning of rain and snow (e.g. ZFL, brightband height, melting layer, and snow elevation) was not addressed in the PVA. Brightband heights, indicative of the altitude at which snowflakes partially melt and transition into rain, offer insights into precipitation processes. Concurrently, the 0-degree isotherm delineates the boundary between freezing and non-freezing temperatures within the atmosphere. To make robust forecast skill assessments, the same physical measurements should be compared. Figure E-5 shows a schematic representation of the different altitudes, and therefore variability, between brightband heights, ZFL, and melting layers. The 0°C isotherm is assumed to be above this layer to compensate for the time/depth of melt to occur and subsequent hydrometer makeup. To utilize profiler observations as a verification source for freezing level forecasts, the difference between the two measurements must be resolved because the observations, models, and physical representation of the rain/snow partitioning may all be different. Most importantly, it affects the level of precision at which one can define freezing level forecast skill.

This section describes the extension of the ZFL forecast assessments of the PVA that address:

- Correct forecasts of freezing conditions at the surface
- The investigation of differences between brightband height and ZFL
- Investigations on the variability of freezing level across complex topography is summarized in Appendix E.1

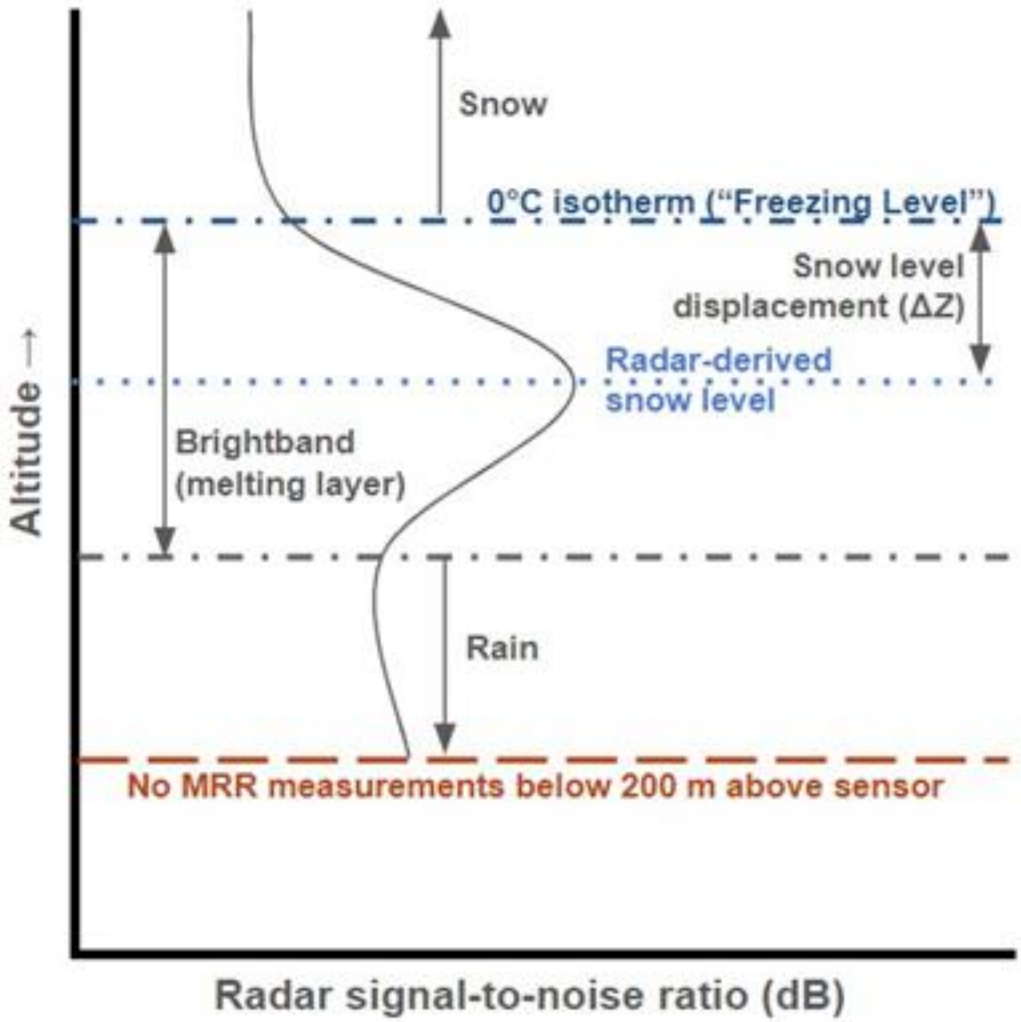


Figure E-5. *Schematic of the height (altitude) 0°C isotherm ("Freezing Level" of ZFL), brightband height ("radar-derived snow level") that is detected by the vertical profiler, melting layer, and the difference between the brightband height and 0°C isotherm (ΔZ). A reference line for missing reflectivity under 200 m is denoted by the black dot-dash line.*

The investigation of correct forecasts of snowing/freezing conditions at the surface was evaluated during the WY2023 season, to leverage several observation types and available high resolution forecast data at the same time -- for the first time possible. We evaluated the skill of the forecasts that accurately predicted freezing conditions when snow was observed. The model forecasts were extracted from CW3E's West-WRF near-real time system, in which forecast predictions are made at 3 different spatial resolutions within the Yuba/Feather region: 9 km, 3 km, and 1 km. In this case, the deterministic forecast with initial and boundary conditions from the GFS model (West-WRF/GFS) are evaluated. The 1 km data was produced for the first time during WY2023. The ZFL forecasts were compared to grid cell elevation within West-WRF to determine whether freezing conditions were observed at the surface. A buffer of 200 m was used to account for the difference between freezing level height and the translation of fall speed and altitude of melting hydrometeors. Disdrometer data, from which precipitation phase can be derived from the distributions of drop size and fall velocity, was used to identify times of snowing conditions as a source for verification. This analysis was conducted for two locations: Downieville (DLA, 901 m/2956 ft), and Lower Bath House (LBH, 1,680 m/5,512 feet). Forecasts of snowing/freezing conditions at the surface were expressed with contingency table metrics (Contingency Table, 2008, Figure E6).

- A forecast hit required that the West-WRF forecasts predicted snow (>=0.5 mm), the forecasted freezing level <= 200 m above the grid cell elevation, and the distrometer reported snow (>0.5 mm).
- A forecast miss is when the forecast did not predict snow but the distrometer measured snow, or the forecast predicted snow but the forecasted freezing level was > 200 m above the grid surface.
- A forecast false alarm is when the forecast predicted snow, the freezing level forecast is near (<200 m) or below elevation, but the disdrometer did not measure snow.
- A forecast correct negative is when the forecast did not predict snow, the freezing level is above terrain (=>200 m), and the distrometer did not measure snow.

Skill is expressed in terms of the Critical Success Index (CSI=Hits/ (Hits+Misses+False Alarms)) and probability of false alarms (POFA=False Alarms/ (False Alarms + Hits)). CSI and POFA range from 0 to 1, where the best values are 1 for CSI, and 0 for POFA.

		Disdrometer Snow Observed	
		Yes	No
Forecast Freezing Level or Snow	Yes	Hit	False Alarm
	No	Miss	Correct Negative

Figure E-6. *Contingency table scenario for forecasting freezing/snowing conditions at the surface.*

The investigation of the differences between brightband height and ZFL serves as a crucial undertaking in understanding atmospheric conditions during winter storms. The verification team was tasked with scoping a potential method for identifying the difference between the brightband height and ZFL with data specific for the Yuba/Feather region. We used a decade's worth of brightband data from Frequency-Modulated Continuous Wave (FMCW) snow level radar (Johnston et al. 2017) at Oroville (OVL) and Micro-Rain Radar (MRR) at New Bullards Bar (NBB) and compared it to CNRFC's publicly available freezing level observed grid (https://www.cnrfc.noaa.gov/fzlvl_guidance.php). The gridded observations represent ZFL instantaneous values at 6-hourly intervals and are direct derivations of the 0-degree isotherm from the High-Resolution Rapid Refresh (HRRR) model analysis (Pete Fickenscher, *personal communication*). Essentially, this dataset contains the best source (long period of record, high spatial resolution) of 0-degree isotherm ZFL overlapping the periods of available profiler observations. The difference of the ZFL and brightband heights were computed by first pairing the median value of all 10-minute profile observations within +/- 1 hour window of the valid time of the gridded observed freezing level. The mean and standard deviation of all profiler-grid pairs were calculated for each WY between 2013 and 2023 and for all water years collectively.

Correct forecasts of freezing conditions at the surface

West-WRF forecast frequency analysis and skill scores of snowing/freezing conditions at DLA using snow observations from the distrometer are given in E-7. The differing resolutions of the model are important because the terrain elevation resolved in the model affects how representative forecasts are of a single point observation. Snowing conditions at Downieville were infrequent, given the large proportion of correct negatives within the forecast. There were proportionally more false alarms within the forecast that is most noticeable at the 9 and 3 km resolution. CSI values for 9 km forecasts across all lead times are only ~0.3, or 30% success ratio, and the probability of false alarms is ~20%. CSI increases slightly as resolution increases (from 9 to 1 km) using lead times of 24-45 hours. POFA also decreases. These two metrics together indicate some benefits of high-resolution forecasts. However, these findings suggest that correctly forecasting snowing events (with subzero temperatures at the surface) remains challenging, given the relatively low CSI. Similar results were seen at LBH.

It is important to note that the grid cell elevation of the model terrain is often not the exact same as the altitude of the disdrometer. However, the goal for this analysis was to determine if the model's elevation/snow compatibility was representative of surface conditions at the point location. We found that the results can be sensitive to the sign of the difference between the grid cell elevation and the observation elevation -- higher grid cell elevations can be disproportionately "more correct" because higher freezing levels are more frequent. This analysis suggests more in-depth studies are needed to affirm resolution impacts on freezing level forecast skill.

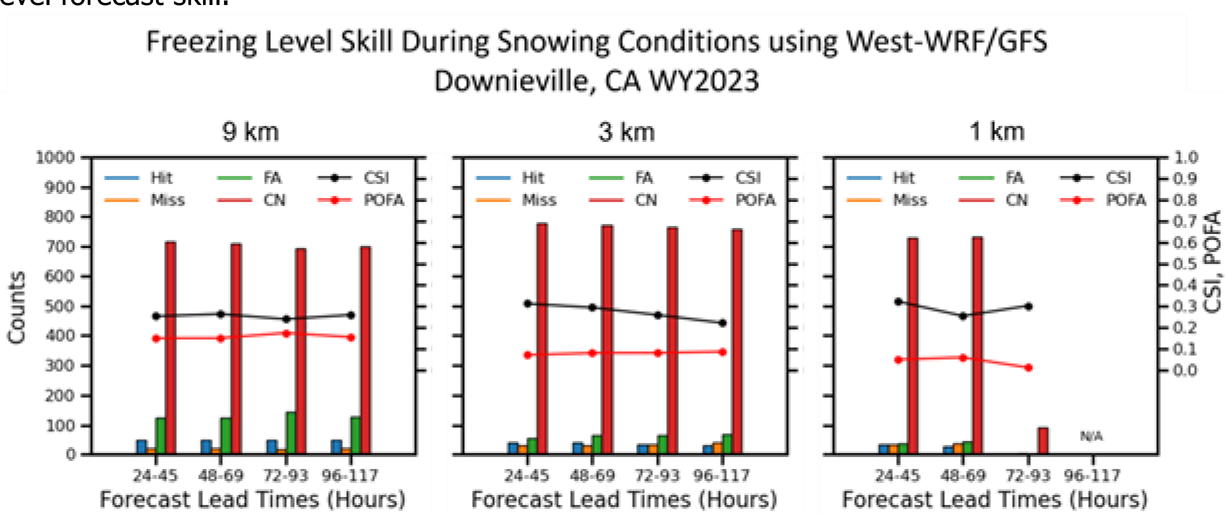


Figure E-7. *Forecast skill of freezing level at Downieville, CA during WY2023: Number of instances (counts) in which West-WRF (WWRF) forecasts are considered hits (blue bars), misses (orange bars), false alarms (green bars), and correct negatives (red bars), along with the forecast skill values of the CSI (black line) and POFA (red line) as a function of lead times using the 9 km (left), 3 km (middle), and 1 km (right) forecasts. All 3-hourly forecast–observation pairs within the e.g. 24–45-hour lead time are evaluated together. The 1 km forecast only extends out to 72 hours.*

The investigation of differences between brightband height and ZFL

The goal of this assessment is to quantitatively determine a value of the offset (Δz) between ZFL and brightband heights and ultimately remove it from either the observation or forecast for a more robust comparison. The mean Δz will indicate average offset between the ZFL and brightband height, and the standard deviation of Δz will indicate whether the mean is representative of most data. In other words, if the standard deviation is small, then the mean value is an accurate representation of the offset. The mean and standard deviation of Δz at OVL are given in E8. Over all years examined, the mean Δz is -187.5 m, and the standard deviation 364.5 m. This translates to a condition in which the actual offset can be as small as 0 m or as large as 552 m. This last value largely exceeds the value used in Sumargo et. al -- meaning, the difference between ZFL and brightband height could represent an even larger proportion of uncertainty, and therefore larger proportion of flood pool space) for runoff events.

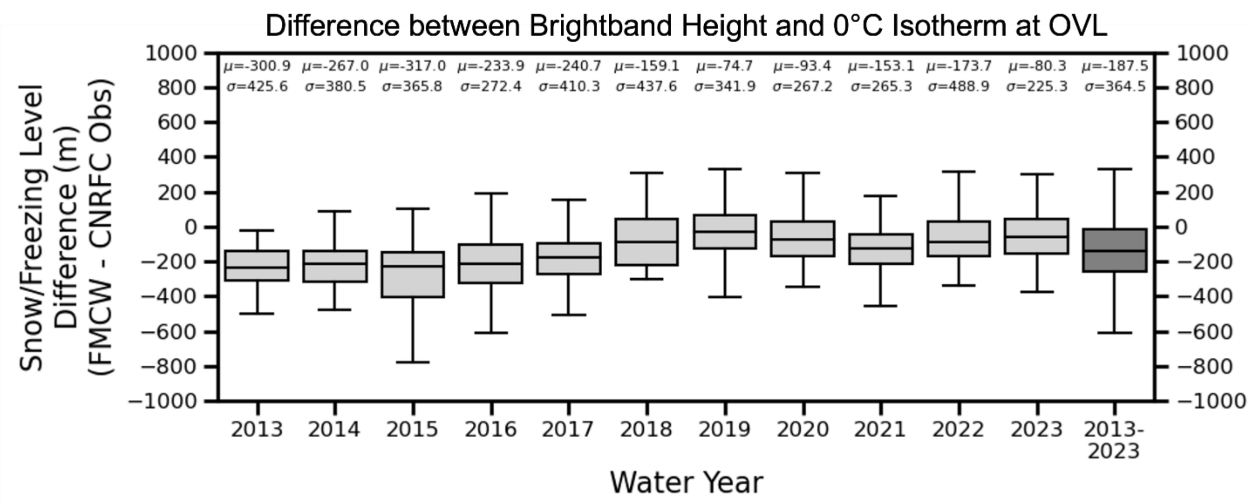


Figure E-8*. Box plots of the differences (Δz, m) between the ZFL and brightband height from the FMCW at OVL as a function of water years between 2013-2023 (all water years are summarized in the last box plot). The mean (**μ**) and standard deviation (**σ**) of Δz are listed above each box plot. All data values are plotted as black dots, the edges of the box correspond to the interquartiles (25th, 75th percentile), and the whiskers correspond to the 5th and 95th percentiles of the data distribution.*

The investigation of freezing level variability over complex terrain

The investigation aimed to assess the variability of freezing levels over complex topography by examining forecast data from West-WRF models at different spatial resolutions. A transect was created, extending from the Central Valley to the Sierra Nevada Crest, with a specific focus on New Bullards Bar. This transect served as the primary sampling path, capturing a range of elevations and topographic features. Forecast data at three different spatial resolutions—1 km, 3 km, and 9 km—were interpolated along the transect to understand how the resolution of the model affects the depiction of freezing level variability, particularly in relation to the orographic influence of the Sierra Nevada mountains. The time frame of interest coincided with the landfall of an atmospheric river (AR) event at 00Z on January 1, 2023, which provided a critical weather context for this analysis. Additionally, two distinct sets of initial conditions were utilized for the West-WRF model runs: one derived from the Global Forecast System (GFS) and the other from the European Centre for Medium-Range Weather Forecasts (ECMWF). This enabled an evaluation of how differing initial conditions influence the predicted freezing levels in the study region.

At the time of the AR landfall, the atmospheric conditions were expected to produce low elevations of the freezing level that intersected with the topography. Figure E-9 shows the transect of the different West-WRF forecasts going from West to East. The different model grid elevations are also given for reference to show the differences in complex topography. The forecasts using different resolutions only differed substantially over the higher topography – where higher resolution forecasts were able to pick up on smaller scale variability of the mountain peaks. Otherwise, the forecasts of differing resolution were identical. The largest forecast variances were due to the different initial conditions: The GFS-based forecasts showed freezing level lower by as much as 500m. There are also signs of possible freezing level bending associated with the approach of the precipitation along the gradient of topography and the presence of precipitation (blue/green bars from the vertical).

This research highlights the sensitivity of atmospheric processes to both model topographic resolution and the selection of initial conditions, emphasizing the need for high-resolution forecasts and robust initial condition strategies to improve understanding of freezing level variability in regions affected by atmospheric rivers.

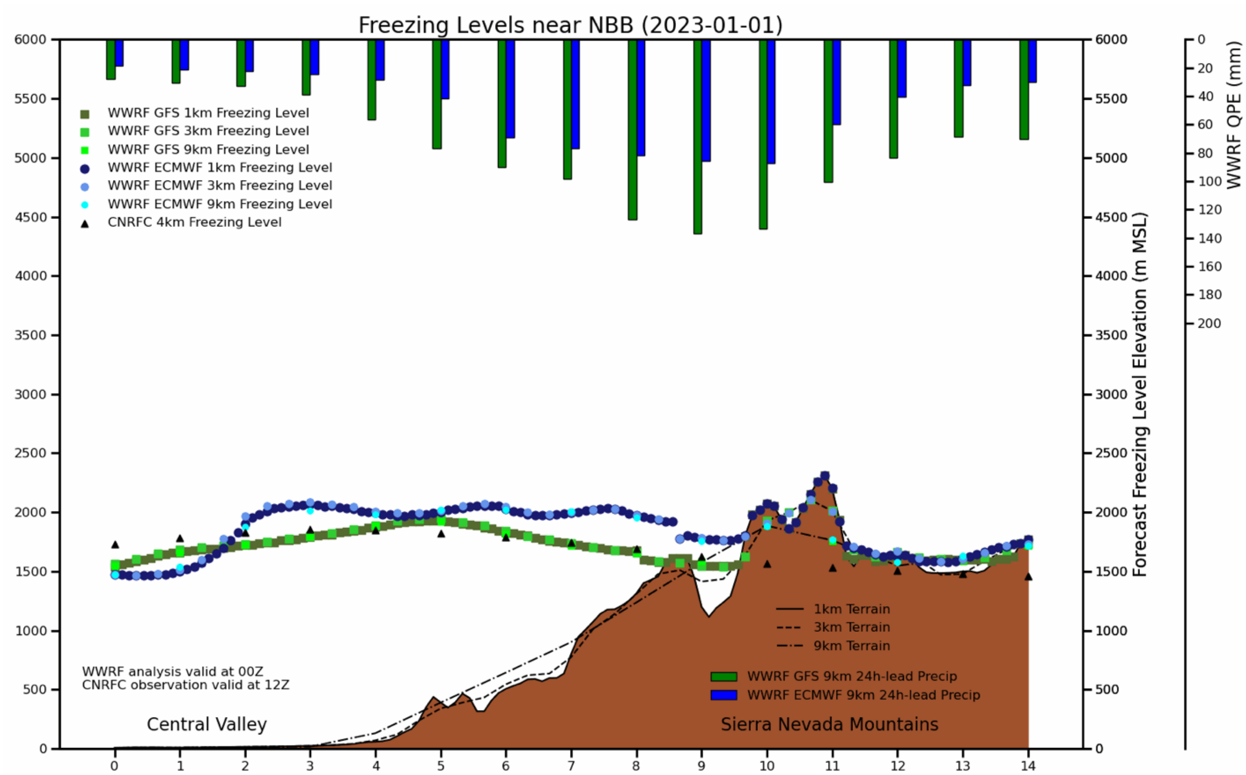


Figure E-9. *Transect of 24-hr lead time freezing level forecasts (symbols) from the West-WRF model forced by GFS (green colors) and ECMWF (blue colors) initial conditions, spanning from the Central Valley to the Sierra Mountain crest, cutting through New Bullards Bar. The blue (green) hanging vertical bars represent the QPF from the ECMWF (GFS) initial conditions at the time of the forecast, and the different model topographic resolutions are given in the black lines.*

These analyses have several implications and generated recommendations for further research. The large spread between the differences in ZFL and brightband height at Oroville could be the result of several factors including meteorological processes (e.g. isothermal layers), the resolution of the model vertical structure within the source data for the gridded observations (recall this is the HRRR analysis), limited precision of the profiler observations, etc. In an attempt remove noise from the profiler data, the all-year $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ decreased to -137.8 m and 209.5 m, respectively. The resolution of topography and initial conditions are important considerations when evaluating the forecast skill of NWP models.

These important findings support further post-FVA investigations including:

- Using IVT/QPE to contextualize differences, especially because latent heating can bend down the freezing level during intense precipitation
- Compare freezing level from radiosondes
- Identify times/depths of isothermal layers
- Compute differences between hydrometeor concentrations in high resolution model predictions as an alternative methodology for Δz determination.
- Include sensitivity on topographic resolution of NWP forecasts

It also creates an opportunity to address the uncertainty in freezing level estimates and potentially engineer a new rain/snow elevation, such that it gives the CNRFC a more reliable estimate of rain/snow partitioning.

E.3. Inflow Forecast Evaluation

Forecasts of 72-hour inflow into New Bullards Bar and Oroville were previously assessed during the Yuba/Feather PVA. This analysis helped to establish a methodology to evaluate AR-related inflow forecast skill and provide quantitative skill evaluations using the CNRFC hindcasts -- a dataset that mimics to a large degree the hydrologic operational ensemble forecast system (HEFS) -- using meteorological forecast inputs relevant at the time. This section describes the extension of the hydrologic forecast assessments of the PVA that address:

- Potential changes in forecast skill due to new meteorological inputs that have been derived since the publication of the PVA
- The need to expand to additional aggregation times (beyond 72-hours) to support operational timelines
- The need for skill assessments at additional locations relevant in the decision-making process in the Yuba/Feather system
- The performance of the hindcast versus current deterministic forecast information

This work leverages the ensemble hindcasts generated using NWS's HEFS in 2022. For a more thorough description of the HEFS system, please refer to Section 6.2 of the PVA and Section 5.3 of this document. The assessments in the PVA leveraged the hindcasts forced with precipitation and temperature data from the GEFSv10, which was generated in 2015. Since then, GEFSv12 meteorological data were generated during a new period of record (2000-2019) and the hindcasts were subsequently updated. The newer meteorological forcings were generated using updated methodologies including new dynamical cores of the model and better spatial resolution. Inflow forecasts for New Bullards Bar (NBB) and Oroville (ORO) were compared using the two versions of the meteorological forcings-based forecasts.

Skill is determined based on the Brier Score (for continuity with the PVA) and Brier Skill Score. Both reflect the error in probabilistic forecasts of a defined event threshold, with the skill score expressing skill against climatology. Lower brier scores are better, versus higher brier skill scores are better where BSS>0 indicates skill over climatology. Thresholds are defined by 95th and 80th percentiles of the observed 72-hr inflow volumes, respectively, to examine the skill of high flow events during AR conditions only (i.e. AR only).

In coordination with the work conducted in Section 3 (Water Resources Engineering), the evaluations were also conducted for 1-day and 7-day total volumes to match address other operational timescales. Forecasts were also evaluated over longer lead times than provided in the PVA (now out to 13 days lead time).

Potential changes in forecast skill due to new meteorological inputs used in CNRFC hindcasts
Table E1 contains the brier scores of the CNRFC ensemble inflow hindcasts at NBB and ORO generated with GEFSv10 and GEFSv12 meteorological forcings. These comparisons were evaluated only during common overlapping periods (i.e. 2000-2010). Note, an error in the calculation of the brier score in the PVA was corrected, so scores in Table E1 supersede the PVA. Interestingly, the brier scores are lower (better) for 72-hr total volume inflows using the GEFSv10 forecasts than in the GEFSv12 forecasts for all lead times out to 5-7 days. Understanding the performance differences is a complex process with several issues to note. The GEFSv12 meteorological ensemble contains 31-members at production, but only 5 members were archived through the entire period of record (a small exception is that 1 11-member run each week was archived). This is compared to the GEFSv10, which contained 10 ensemble members. The number of hindcast ensemble members is also different between versions. Changes to the MEFP used to drive the ensemble forecasts also occurred, included 1) moving to a new climatology which limited the number of members created from the MEFP distributions, 2) changing the sampling method within MEFP going from stratified random sampling to fixed quantile sampling, 3) recalibrating all hydro models to use freezing level to define the rain/snow line instead of temperature, and 4) different sampling techniques embedded within the MEFP (Brett Whitin, 2023 *personal communication*). There are also considerations about the precipitation skill of the GEFS, which serves as one of the input sources for the hydrologic model calibration.

Despite the differences, both hindcast versions have brier scores near zero which implies high forecast skill during AR events. Several recommendations for further research to address the difference in the GEFS hindcast skill include:

- Comparing reliability diagrams (v10 vs v12) to identify any potential important differences between the two MEFP sets or advantages in capturing larger events (hits versus false alarms)
- Comparing MEFP outputs from the different hindcasts to understand potential sampling bias
- Compare GEFS precipitation skill during the period in which hindcasts were calibrated to understand the impact of precipitation forcing

	Brier Score (AR only)

Lead Time Aggregate	NBB		ORO	
	GEFSv10	GEFSv12	GEFSv10	GEFSv12
1-3 days	**0.063**	0.076	**0.042**	0.051
4-6 days	**0.087**	0.110	**0.070**	0.081
7-9 days	**0.120**	0.140	**0.091**	0.120

Table E-1. *Brier scores of CNRFC ensemble inflow hindcasts for 2000–2010 at NBB and ORO. The scores are computed with an 80th flow percentile threshold, for lead time aggregates of one to three, four to six, and seven to nine days and AR-only conditions. Bold scores indicate the version of GEFS with better skill.*

Brier Skill Score updates for additional lead times and sites using CNRFC hindcasts
Table E-2 contains the Brier Skill Score (BSS, higher is better) for the GEFSv12 as an update to the PVA findings of 72-hour streamflow from the CNRFC hindcasts. This table now includes scores for Englebright (ENG), as well as NBB and Oroville (ORO) out to 13 days lead time. BSS > 0 indicates that the forecasts are skillful beyond climatology. Using the GEFSv12 hindcasts, the 72-hr total volume flows have BSS > 0 out to 7-9 days lead time at all three locations. Additionally, the scores are aggregated to 24-hr total volumes and 168-hr total volumes. 24-hour total volumes have BSS > 0 out to 6 days at all locations and 168-hr total volumes are skillful out to 7-14 days lead time.

Lead Time Aggregate	Brier Skill Score (AR only) GEFSv12		
	NBB	ORO	ENG
	72-hr total volumes		
1-3 days	**0.5493**	**0.6637**	**0.5369**
4-6 days	**0.3343**	**0.4583**	**0.3234**
7-9 days	**0.0779**	**0.1722**	**0.071**
10-13 days	-0.1588	-0.0476	-0.1278
	24-hr total volumes		
1-day	**0.3706**	**0.5493**	**0.4362**
2-day	**0.4087**	**0.5223**	**0.486**
3-day	**0.3077**	**0.4186**	**0.3768**
4-day	**0.1813**	**0.3671**	**0.2839**
5-day	**0.1601**	**0.254**	**0.2194**
6-day	**0.063**	**0.1254**	**0.0972**
7-day	-0.118	-0.0569	-0.0534
8-day	-0.196	-0.1637	-0.1482
9-day	-0.3377	-0.3073	-0.2574
10-day	-0.4336	-0.381	-0.3581
	168-hr total volumes		
1-7 days	**0.064**	**0.053**	**0.068**
8-14 days	**0.160**	**0.120**	**0.130**

Table E-2. *Brier skill scores of CNRFC ensemble inflow hindcasts for 2005-2019 at NBB, ORO, and ENG only during AR conditions. The scores are computed with an 80th flow percentile threshold, for lead time* *aggregates of 72-hours, 24-hours, and 168-hours. Bold scores indicate lead times where skill is better than climatology.*

Performance of the hindcast versus current deterministic forecast information
A comparative analysis was conducted between the operational CNRFC deterministic forecasts and the 75th percent exceedance value of the hindcast for 72-hour total volumes at 1–3-day lead times. It is often noted that the 75th percent exceedance value of the operational ensemble forecast aligns well with the deterministic forecast; therefore, using the 75th percentile value of the hindcast is one way to directly compare the model performance of data used to support water resources engineering research (and the assessments herein) with what is currently used operationally for water management decision support in the Folsom Dam and Lake WCM. Additionally, FIRO alternatives ID3A for both ORO and NBB use 75% NEP for daily volumes up to 7 days (Section 3). In summary, the coefficient of determination (R^2) is higher in the operational forecasts at NBB and ORO and the RMSE is smaller. This suggests that operational forecasts may have better skill than what was assessed herein with the hindcasts. Additionally, performance of the operational deterministic 72-hour inflow volume forecasts for Englebright, Oroville, and New Bullards Bar at the 1–3-day lead time was evaluated. In short, the variance explained by the forecast was 76%, 89%, 86%, respectively. This demonstrates that the forecasts are capturing a large majority of the observed forecast variability.

E.4. Meteorological Ensemble Forecast Evaluation

Meteorological probabilistic forecasts play an important role in providing uncertainty estimates amongst single forecast predictions. Ensembles provide a range of possible outcomes that can be filtered to express likelihoods of precipitation, IVT, etc. The PVA assessed several different deterministic (or ensemble mean) forecast products to convey skillful lead times of extreme precipitation. This section describes the extension of meteorological forecast skill assessments in the PVA with additional evaluations of ensemble forecasts of precipitation and landfalling IVT.

CW3E has been producing a 200-member meteorological ensemble beginning in WY2022 and is summarized in fine detail in the PVA Appendix M.3. Updates for WY2023 on the model system are already mentioned in section 5.2.3.e. For brevity, the model configuration details are not repeated here. However, this section focuses on new analysis of the model performance of probabilistic precipitation during the December 2021-March 2022 winter season. The assessment also includes a comparison to the Deep Learning method applied to the 200-member ensemble also discussed in section 5.2.3.e. Precipitation is evaluated against the Parameter-elevation Relationships on Independent Slopes Model (PRISM, https://prism.oregonstate.edu/) 4 km dataset, which serves as the observation dataset and daily climatology.

To understand the probabilistic skill of IVT at landfall, the Global Ensemble Forecast System (GEFS) was examined using the "AR landfall tool" (Cordeira and Ralph 2021) introduced in section 5.2.3.d. This section discusses the qualitative forecast evaluation of the AR landfall tool for a series of events occurring during a period of successive AR activity during WY2023 (referred to as the Deep Dive period). The ensemble probabilities are calculated for a given location and aligned with observed IVT to determine how well the ensemble probabilities reflected observed AR conditions as a function of forecast lead time. The ensemble tool explicitly reflects the number of ensemble members that are predicting IVT greater than a defined threshold.

Forecast skill is assessed qualitatively by comparing the alignment of observed precipitation (IVT) with probabilities of precipitation (IVT) using specific thresholds. For more quantitative measures, the Continuous Ranked Probability Skill Score (CRPSS) and Brier Skill Score (BSS) are used to define skill for precipitation. CRPSS quantifies the overall difference between the forecasted probabilities and the actual distribution of events. Higher values of CRPSS are better, where a value of 0 means the forecasts are no better than climatology. BSS is the mean squared error relative to climatology, where a value of 1 is best and a value of 0 means the forecasts are no better than climatology.

Performance of the 200-member ensemble

Figure E-10 shows the probabilities of 24-hour accumulated precipitation > 1 inch (25.4 mm) for a representative test case (24 December 2021) from GEFS (Figure E-10A), West-WRF 200-member ensemble (Figure E-10B), and West-WRF 200-member ensemble + Deep learning (Figure E-10C). The application of the deep learning technique is shown to improve the skill of the raw ensemble forecast. The larger probabilities generally align well with the observed precipitation of 1 inch or greater (black contours) across all models, but the West-WRF 200-member ensemble and that with Deep Learning applied have greater probabilities than the GEFS over the Sierras -- implying more certainty that precipitation will exceed 1 inch or more in these areas.

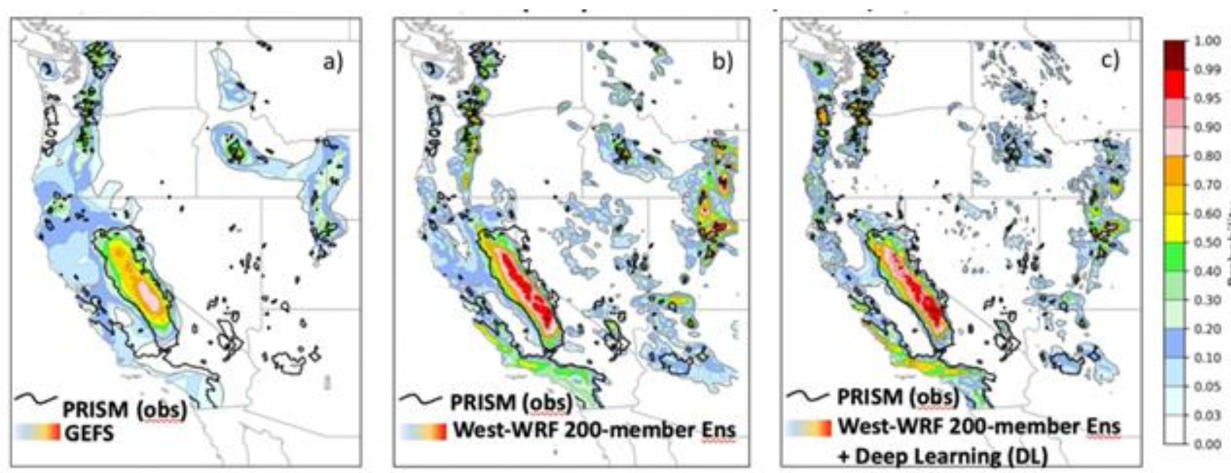


Figure E-10. *Probability comparison among GEFS (a), ECMWF, West-WRF 200-mem ensemble, and West-WRF 200-mem ensemble + deep learning of 24-hr precipitation > 1 inch (25.4 mm) valid 24 Dec 2021. The dark black line represents the observed 1-inch (25.4 mm) precipitation contour from PRISM.*

Furthermore, for the Yuba-Feather watershed, the deep learning post-processed West-WRF forecast (shown in purple; E-11 left) outperforms (i.e., CRPSS is larger) all other reference benchmark models, including the ECMWF (green) and the raw West-WRF (red), for the period of assessment (Dec 2021-Mar 2022), from a lead time of 1 to 3 days. The raw 200-member ensemble is competitive with the ECMWF and GEFS predictions, although it has an equivalent or larger CRPSS at all lead times compared to the GEFS. The CRPSS represents skill across the spectrum of all precipitation thresholds, the brier skill score is calculated for specific thresholds.

E-11 (right) shows that the deep learning-applied approach has a clearer improvement over the raw ensemble counterparts particularly for this upper threshold of 50 mm (1.96 inches) out through 4 days lead time. This suggests that the large West-WRF ensemble plus the post-processing has desirable added value for predictions of precipitation in the West.

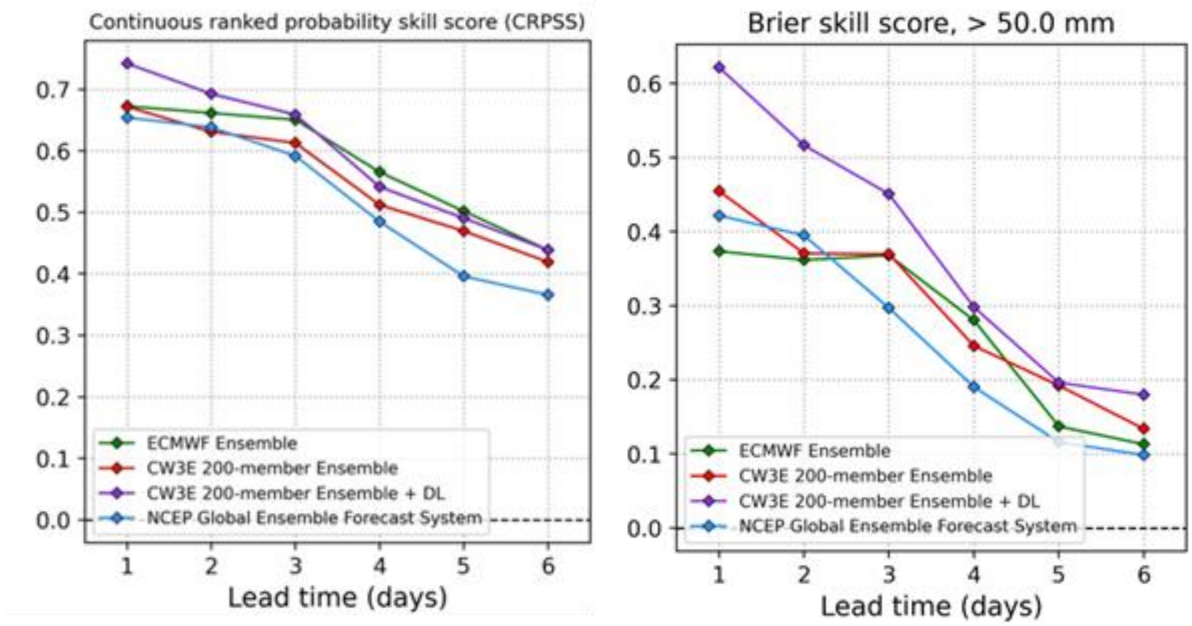


Figure E-11. *(Left) Continuous ranked probability skill score (CRPSS) and (right) brier skill score among GEFS (blue), ECMWF (green), West-WRF 200-mem ensemble (red), and West-WRF 200-mem ensemble + deep learning of 24-hr precipitation (purple) using combined winter season forecasts during WY2022 and WY2023. The CRPSS is calculated over the entire domain and reflects skill for all precipitation thresholds, and the brier score is calculated for a precipitation threshold of > 50 mm (1.96 in)*

Performance of the GEFS ensemble probability of IVT in Northern California

Section 5.2.3.d introduces the "AR landfall tool" as a method for examining the probability of IVT exceeding defined thresholds at a certain location along the U.S. West Coast. Figure E-12 shows the probability of IVT > 250 kg m^{-1} s^{-1} at a point along the coast near Bodega Bay (37.5N, 122.5W) for the sequence of ARs making landfall between 17 December 2022 and 16 January 2023 (herein Deep Dive Period) and the actual IVT magnitude as observed from the GEFS analysis. Approximately 8 ARs made landfall in this region during the Deep Dive Period, as signified by the shaded red areas in the observed IVT time series, and the areas in purple represent where more than 90% of ensemble members were predicting IVT > 250 kg m^{-1} s^{-1}. Using a percentage threshold of 50% (i.e. more than 50% of ensemble members) as one example of forecast skill, the GEFS ensemble predicted landfalling IVT> 250 kg m^{-1} s^{-1} anywhere between 6-13 days ahead of time. If that threshold is increased to 75% of the ensemble members, IVT > 250 kg m^{-1} s^{-1} was correctly predicted from 5-12 days ahead of time. There is clear variability in the lead time predictability from storm to storm across the deep dive period. This type of analysis is helpful in that it can guide further research into understanding meteorological patterns associated with specific storms that may point to explanations of lead time predictability (e.g. blocking patterns). Section 5.2.3.d also found over a longer period of record that GEFS was able to predict IVT 250 kg m^{-1} s^{-1} in Northern California up to 6 days, on average, for moderate ARs on the AR scale (i.e. AR2). More detail is provided in Appendix B.2.4.

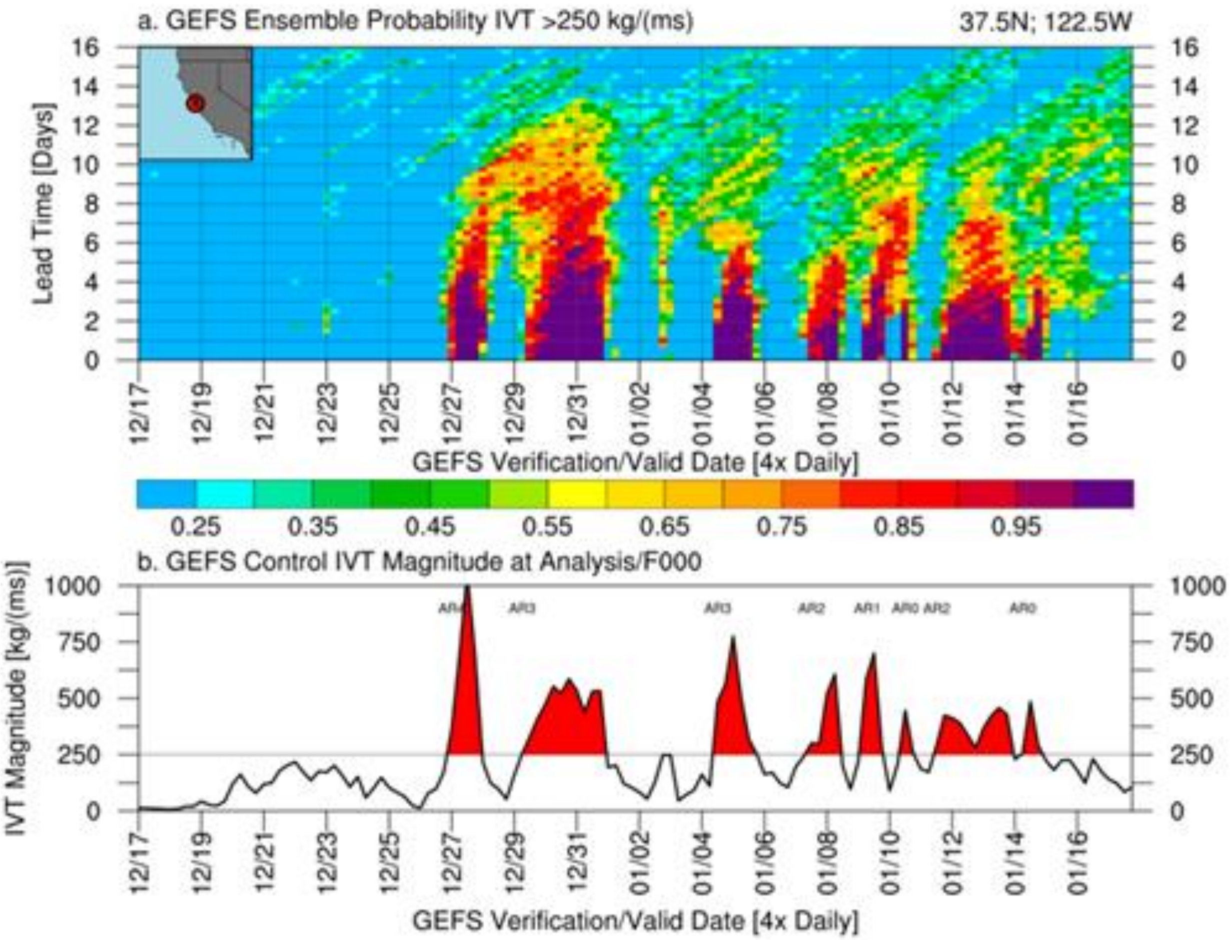


Figure E-12. *(a) A verification time – lead-time analysis of the ensemble odds of IVT magnitudes >=250 kg m^{-1} s^{-1} (shaded) from the NCEP GEFS for forecasts verifying between 17 December 2022 through 17 January 2023. (b) A time series of IVT magnitude from the GEFS control IVT magnitude (kg m^{-1} s^{-1}) at the 0-hour forecast time (e.g. analysis) with periods with IVT magnitudes >=250 kg m^{-1} s^{-1} shaded in red.*

E5. References

Contingency Table (2008). In: The Concise Encyclopedia of Statistics. Springer, New York, NY. https://doi.org/10.1007/978-0-387-32833-1_77

Cordeira, J. M., and F. M. Ralph, 2021: A Summary of GFS Ensemble Integrated Water Vapor Transport Forecasts and Skill along the U.S. West Coast during Water Years 2017–20. Wea. Forecasting, 36, 361–377, https://doi.org/10.1175/WAF-D-20-0121.1.

Johnston, P. E., Jordan, J. R., White, A. B., Carter, D. A., Costa, D. M., & Ayers, T. E. (2017). The NOAA FM-CW snow-level radar. Journal of Atmospheric and Oceanic Technology, 34(2), 249–267. https://doi.org/10.1175/JTECH-D-16-0063.1

Sumargo, E., Cannon, F., Ralph, F. M., & Henn, B. (2020). Freezing level forecast error can consume reservoir flood control storage: Potentials for Lake Oroville and New Bullards Bar

reservoirs in California. Water Resources Research, 56, e2020WR027072.
https://doi.org/10.1029/2020WR027072