Title

Advancing Hydrological Forecasting and Reservoir Management Through Model Calibration and SWOT Satellite Observation

ISBN

9798190935974

Author

Ma, Xiaoyu

Publication Date

2026-09-11

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Advancing Hydrological Forecasting and Reservoir Management Through Model Calibration and SWOT Satellite Observation

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in Geography

by

Xiaoyu Ma

2026

ABSTRACT OF THE DISSERTATION

Advancing Hydrological Forecasting and Reservoir Management Through Model Calibration and SWOT Satellite Observation

by

Xiaoyu Ma

Doctor of Philosophy in Geography

University of California, Los Angeles, 2026

Professor Dennis P. Lettenmaier, Committee Chair

Hydrological prediction depends on both accurate models of system dynamics and observations that constrain hydrologic model predictions. This dissertation investigates how advances in hydrologic modeling and satellite remote sensing can improve forecasting and support water-management decisions across data-rich and data-limited regions.

First, I evaluate short-range flood forecasting using the Noah-MP land-surface model (which is the hydrologic core of the U.S. National Water Model) in three California watersheds spanning contrasting hydroclimatic conditions from north to south. Model calibration is well known to be essential for accurate flood forecasts, and I confirmed that careful calibration substantially improved retrospective forecasts (reforecasts) of flood hydrographs and peak flows. In addition, correction of precipitation forecast biases and adjustment of antecedent soil moisture through discharge-informed data assimilation further improved reforecast skill, particularly beyond two-day lead times in the southernmost semi-arid basin among Russian River Basin, Yuba-Feather Basin, and Santa Ana Basin. The resulting Noah-MP forecasts were comparable to, and in several metrics slightly more accurate than, archived operational forecasts provided by the U.S. National Weather Service California-Nevada River Forecast Center.

The second part of this dissertation focuses on reservoir monitoring and forecasting using observations from the Surface Water and Ocean Topography (SWOT) satellite mission, a cooperative mission of the U.S. and France launched in late 2022. Across 12 western U.S. reservoirs, SWOT water-surface elevations had a median absolute error below 20 cm, while derived storage errors were generally below 5% compared with in-situ observations. Initializing SWOT-derived storage into a reservoir-operation model enabled temporally continuous storage estimates between satellite overpasses. I then integrated SWOT-derived initial storage, seasonal inflow forecasts, and reservoir-operation modeling into a seasonal forecasting framework. Evaluation at two California reservoirs, New Bullards Bar reservoir and Don Pedro reservoir, showed that storage initialization is most influential at short lead times, whereas inflow uncertainty increasingly dominates at longer leads. Application to Tendaho Reservoir, Ethiopia, demonstrated the potential of combining satellite observations and globally available forecasts to assess seasonal water availability in a setting where continuous in situ observations are limited.

Together, these studies show that integrating improved models and observations can advance forecast-informed water management across multiple timescales and observational settings.

The dissertation of Xiaoyu Ma is approved.

A. Park Williams

Yongwei Sheng

Rong Fu

Steve Margulis

Dennis P. Lettenmaier, Committee Chair

University of California, Los Angeles

2026

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION ii

LIST OF FIGURES x

LIST OF TABLES xii

ACKNOWLEDGEMENTS xiii

VITA xv

Chapter 1 Introduction 1

  References 6

Chapter 2 Model Calibration and Data Assimilation for the National Water Model: A Case Study over California 12

  Abstract 12

  2.1 Introduction 13

  2.2 Study Description and Methods 17

    2.2.1 Study Area and Forecast Points 17

    2.2.2 Land Surface Model and Data Sets 20

      2.2.2.1 Hydrology Model 20

      2.2.2.2 Meteorological Forcings 21

  2.3 Experimental Design 22

    2.3.1 Model Parameter Estimation 22

2.3.2 Historical Flood Reconstruction ... 24

2.3.3 Bias Correction of QPF ... 25

2.3.4 Adjustment of Antecedent Soil Moisture through Data Assimilation ... 26

2.4 Results ... 29

2.4.1 Streamflow Simulation after Model Calibration ... 29

2.4.2 Flood Reforecasts after QPF and Soil Moisture Adjustments ... 31

a. Peak Magnitude Differences ... 33

b. Peak Time Differences ... 36

c. Flood Hydrograph Shape Differences ... 37

2.5 Discussion ... 39

2.6 Conclusions ... 41

Acknowledgements ... 42

References ... 43

Chapter 3 Monitoring Reservoir Storage Using SWOT Satellite Observations and a Reservoir Operation Model ... 48

Abstract ... 48

3.1 Introduction ... 49

3.2 Data and Methodology ... 52

3.2.1 Study Area ... 52

3.2.2 SWOT WSE and SA Observations ... 55

3.2.3 In Situ WSE, Storage, and Inflow Observations 57
3.2.4 Derived In Situ-Based SA Estimates 58
3.2.5 Estimation of Reservoir Storage 58
3.2.5.1 SWOT-Based Reservoir Storage Estimates 58
3.2.5.2 Model-Based Reservoir Storage Estimates (Hanasaki Algorithm) 61
3.2.5.3 SWOT Gap-Filling Using the Hanasaki Algorithm 63
3.3 Results 64
3.3.1 SWOT WSE Observations 64
3.3.2 SWOT SA Observations 67
3.3.3 SWOT Storage Estimates 71
3.3.4 Hanasaki Storage Estimates 75
3.4 Discussion 80
3.5 Conclusions 82
Acknowledgments 82
References 84
Chapter 4 SWOT-Initialized Reservoir Operation Forecasting: Validation in Data-Rich and Application to Data-Limited Regions 91
Abstract 91
4.1 Introduction 92
4.2 Study Areas 95

4.2.1 Data-Rich Validation Reservoirs: BUL and DNP 95

4.2.2 Data-Limited Application: Tendaho Reservoir 96

4.3 Data 97

4.3.1 SWOT LakeSP Observations 98

4.3.2 In Situ Reservoir Observations 99

4.3.3 Seasonal Inflow Forecasts 99

4.3.4 Reservoir Attributes and Agricultural Water-Demand Data 100

4.4 Methods 101

4.4.1 Forecasting Framework and Experiment Design 101

4.4.2 Seasonal Reservoir Inflow Forecasts 102

4.4.3 SWOT-Based Reservoir-Storage Initialization 105

4.4.4 Hanasaki Reservoir Operation Model 106

4.4.5 Forecast Evaluation and Agricultural Demand-Satisfaction Assessment 107

4.5 Results 108

4.5.1 SWOT-Based Reservoir Storage Initialization 108

4.5.2 Seasonal Reservoir Inflow Forecasts 110

4.5.3 Reservoir Storage Forecasts 112

4.5.4 Forecasted Agricultural Water Supply Relative to Estimated Demand 118

4.5.5 Application of the Validated Workflow in a Data-Limited Region: Tendaho Reservoir. 120

4.6 Conclusions 123

References 125

Chapter 5 Conclusion 134

Appendix A Supplementary Material for Chapter 2 139

Table S1. POT3 Events for the Six Selected FIRO Points 140

Figure S1. Weighted Average Flood Hydrographs for Observations and Noah-MP Simulations 142

Appendix B Supplementary Material for Chapter 4 143

Section S1. Evaluation of Historical Inflow Simulations and GloFAS Bias Correction 143

Section S2. Evaluation of Inflow Forecasts Initialized during the Low-Flow Period 146

Section S3. SWOT-Observed Water Surface Elevation and Derived Storage at Tendaho Reservoir 148

LIST OF FIGURES

Figure 2.1 Map of study region showing the three FIRO watersheds. 28

Figure 2.2 Normalized precipitation hydrograph weighted by QPE peaks, using Yuba-Feather DCSC1 as an example to illustrate the systematic bias in QPF relative to QPE. 29

Figure 2.3 Scatter plot of simulated flood streamflow and observed flood streamflow in seven forecasting points in the FIRO watersheds. 30

Figure 2.4 KGE for Russian (a), Yuba-Feather (b), and Santa-Ana (c) Rivers for lead times 1, 2, 3, and 4 days. CNRFC forecast results are shown in blue; Noah-MP reforecasts in yellow line; and Noah-MP reforecasts after QPF bias correction and data assimilation are shown in green. 32

Figure 2.5 Median (solid lines) and interquartile range (shading) of the relative differences of floods peak (model minus observed divided by observed) streamflow of Noah-MP reforecasts and RFC forecasts as a function of forecast hours in advance of the observed peak (lead time) in Russian (a), Yuba-Feather (b), and Santa-Ana (c) for lead times for 1, 2, 3, and 4 days. 34

Figure 2.6 Boxplots of flood peak time differences versus lead days (1, 2, 3, and 4 days) for CNRFC forecasts (blue box), Noah-MP unadjusted forecasts (yellow box), and Noah-MP forecasts after QPF bias and initial condition adjustments (green box) in (a) Russian, (b) Yuba-Feather, and (c) Santa Ana River Basins. 36

Figure 2.7 Normalized Flood Hydrographs (Weighted by Flood Peaks) for Russian (a-c), Yuba-Feather (d-f), and Santa-Ana (g-i) Rivers. The flood hydrographs are historical forecasts from CNRFC (first column), Noah-MP unadjusted (second column), and Noah-MP adjustment for QPF bias and soil moisture initial conditions (third column) at different lead days. 38

Figure 3.1 Locations of the 12 selected reservoirs in the Western U.S. Circle sizes represent the storage capacity of each reservoir (see inset). 53

Figure 3.2 (a) Schematic illustration of baseline WSE and storage. (b) In situ WSE-storage curves, and corresponding fitted functions for the 12 WUS reservoirs in California (first column), the UCRB (second column), and CRB (third column).. 60

Figure 3.3 Time series of SWOT WSE against in situ observations in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column)......... 65

Figure 3.4 Derived in situ WSE-SA curves compared with SWOT-derived WSE-SA pairs over the WUS in California (first column), the UCRB (second column), and CRB (third column)... 69

Figure 3.5 Time series of SWOT SA against derived in situ SA in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column)....................... 71

Figure 3.6 Time series of SWOT reservoir storage against in situ storage observations in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column). 72

Figure 3.7 Hanasaki-based storage estimates compared with in situ storage observations from 2009 to 2025 for the 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column). 76

Figure 3.8 Time series of Hanasaki reservoir storage estimates when there are no (upper row), less than a quarter (second row), half (third row), and full (bottom row) SWOT observations available in DNP, NAV, and RIM. 78

Figure 4.1 Study reservoirs and associated SWOT LakeSP (lake single pass) polygons: (a) New Bullards Bar Reservoir, California, USA; (b) Don Pedro Reservoir, California, USA; and (c) Tendaho Reservoir, Ethiopia. Symbols indicate the approximate dam/outlet location, reservoir inflow location, and reservoir center. Insets show the regional location of each reservoir. 97

Figure 4.2 Conceptual framework for validating and transferring a SWOT-informed reservoir forecasting workflow from data-rich to data-limited regions. 102

Figure 4.3 Comparison of SWOT-derived and in-situ reservoir storage and elevation-storage relationships for New Bullards Bar and Don Pedro reservoirs. Panels (a) and (b) show observed storage, storage derived directly from SWOT observations, and storage estimated by combining SWOT and in-situ information for New Bullards Bar and Don Pedro, respectively. Panels (c) and (d) compare the water-surface elevation-storage relationships for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP), respectively, using agency records, SWOT-derived observations, and combined SWOT-in-situ reference points. 108

Figure 4.4 Seasonal reservoir inflow forecasts from ESP and GloFAS for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP) across monthly forecast issue dates. 111

Figure 4.5 Comparison of observed and SWOT-based reservoir storage at monthly forecast issue dates for New Bullards Bar Reservoir (a; BUL) and Don Pedro Reservoir (b; DNP). Blue circles indicate observed (in situ) storage at the forecast initialization date (T0), red crosses indicate the latest available SWOT-derived storage observations before T0, black dashed lines show their propagation to T0, and orange circles indicate the resulting propagated SWOT-based storage estimates at T0. 112

Figure 4.6 Comparison of ESP- and GloFAS-driven reservoir storage forecasts initialized with SWOT-based storage in comparison with observed (in situ) storage for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP) in 2024 and 2025. Black solid lines with circles represent observed storage, orange dashed lines represent forecasts driven by ESP inflows, green dashed lines represent forecasts driven by GloFAS seasonal inflows, green shading indicates the GloFAS forecast uncertainty range, and red crosses identify the SWOT-based initial storage at each forecast issue date. Each panel corresponds to an individual monthly forecast issue date. 114

Figure 4.7 Monthly reservoir storage forecast errors under SWOT-based and observed-storage initialization for BUL and DNP using ESP and GloFAS inflow forecasts in 2024 and 2025... 116

Figure 4.8 May-August agricultural water-supply derived from monthly initialized reservoir simulations forced by ESP and GloFAS inflow forecasts, compared with irrigation-demand estimates for other (all) years and dry years. Dry years are defined as the five driest water years during the agriculture record period (water year 1980-2020 for BUL; year 2012-2023 for DNP). 118

Figure 4.9 GloFAS-driven seasonal reservoir forecasts for Tendaho Reservoir. The upper panels show forecast storage trajectories initialized with SWOT-derived storage at T0 for eight issue months in 2024 and 2025, while the lower panel compares monthly mean forecast release with estimated irrigation demand. 121

Figure A1. Weighted average flood hydrographs for observations (red) and Noah-MP simulations (blue).. 142
Figure B1. Evaluation of bias-corrected GloFAS and Noah-MP-based historical monthly inflow simulations relative to in situ observations for New Bullards Bar Reservoir (BUL), California. The time series compares observed monthly mean inflow with raw GloFAS, bias-corrected GloFAS, and Noah-MP simulations during 2007–2023. The scatterplot compares monthly inflow simulated by bias-corrected GloFAS and Noah-MP against in situ observations. The dashed diagonal line represents the 1:1 relationship.. 144
Figure B2. Evaluation of bias-corrected GloFAS and Noah-MP-based historical monthly inflow simulations relative to in situ observations for Don Pedro Reservoir (DNP)........................... 145
Figure B3. Comparison of ESP and GloFAS monthly inflow forecasts initialized during the May–August low-flow period at New Bullards Bar Reservoir (BUL) for 2024 and 2025. 147
Figure B4. Comparison of ESP and GloFAS monthly inflow forecasts initialized during the May–August low-flow period at Don Pedro (DNP) for 2024 and 2025.................................. 147
Figure B5. Time series of SWOT-observed water surface elevation and derived storage at Tendaho Reservoir (TDH), Ethiopia... 149

LIST OF TABLES

Table 2.1 Drainage area, observing sites, and RFC sites for the study region. 19
Table 3.1 Names, Sizes, Regions, Main Usages, and Sources of In Situ Observations of the 12 Selected Reservoirs in the WUS (California Includes the Sacramento and San Joaquin River Basins, Columbia Is the Columbia River Basin Above the Dalles, OR, and UCRB Is the Upper Colorado River Basin Above Lees Ferry, AZ). 54
Table 3.2 MAE and 90th PAE of SWOT WSE Observations in the 12 Reservoirs................... 66
Table 3.3 Statistical Performance (R2 and KGE) Comparing Simulated and Observed Reservoir Storages.. 77
Table A1 POT3 events for the six selected FIRO points... 140

ACKNOWLEDGEMENTS

I would first like to express my deepest gratitude to my advisor, Professor Dennis P. Lettenmaier, for his guidance, support, and encouragement throughout my doctoral studies at UCLA. His scientific insight, high standards, and dedication to research have profoundly shaped the way I think about hydrology and scientific inquiry. I am especially grateful for the trust and independence he has given me to explore different research directions throughout my Ph.D. journey.

I would also like to thank my dissertation committee members, Professor Park Williams, Professor Rong Fu, Professor Yongwei Sheng, and Professor Steve Margulis, for their thoughtful feedback, insightful questions, and support of my dissertation research.

I am deeply grateful to Professor Jeff Dozier, Professor Yongkang Xue, and Professor Ming Pan for their guidance at different stages of my doctoral training. I am especially grateful to the late Professor Jeff Dozier, whose intellectual generosity and encouragement left a lasting impression on me.

I have been fortunate to receive advice, collaboration, and friendship from many colleagues and labmates. I would especially like to thank Zhaoxin Ban, Lu Su, Mu Xiao, Qian Cao, Dongyue Li, Yiwen Fang, Zhijiong Cao, Khalil Ganem, Haorui Sun, Yihan Wang, and Qian He for their generous help, thoughtful discussions, and support over the years.

I am also grateful to the staff in the UCLA Department of Geography, especially Kasi McMurray, Jenee, Lukas, and Brian, for their patience, kindness, and assistance throughout my graduate studies.

Outside of research, I am deeply thankful to my friends Yanyan Li, Yongjia He, and Kiki Qin for their encouragement and for sharing many moments of happiness and laughter throughout my Ph.D. years.

I owe special thanks to my husband, Ziheng Zhang, whose optimism, encouragement, and unwavering support have carried me through many difficult times. I am also grateful to my cat, Moon, who kept me company through many late nights of writing, coding, and working. Finally, I would like to thank my parents for their unconditional love, sacrifices, and support throughout every stage of my life.

I gratefully acknowledge the collaborators, institutions, and funding sources that supported this dissertation. This work was supported by the National Aeronautics and Space Administration through NASA SWOT Science Team Grant 80NSSC20K1340 to the University of California, Los Angeles, and by the UCLA Graduate Division through the Dissertation Year Award. Chapters 2–4 are based on work for which I was the primary researcher and lead author. Chapter 2 was published in the *Journal of Hydrometeorology*, Chapter 3 in *Water Resources Research*, and Chapter 4 is currently under review at *Hydrology and Earth System Sciences*. I am grateful to my coauthors and research collaborators for their contributions and collaboration.

VITA

EDUCATION

Ph.D. Candidate in Geography, University of California, Los Angeles	2022
M.S. Civil and Environmental Engineering, University of California, Los Angeles	2020
B.S. Hydrology and Water Resources, Hohai University	2019

PROFESSIONAL EXPERIENCES

Graduate Researcher: University of California, Los Angeles	2020–current
Teaching Assistant: University of California, Los Angeles	Mar–Jun 2025

SELECTED PUBLICATIONS

Ma, X., and D. P. Lettenmaier, 2026: SWOT-initialized reservoir operation forecasting: Validation in data-rich and application to data-limited regions. Hydrology and Earth System Sciences, under review.

Ma, X., A. Park Williams, and D. P. Lettenmaier, 2026: Is the global hydrological cycle accelerating? An analysis based on global precipitation trends. Proceedings of the National Academy of Sciences, under review.

Ma, X., R. K. Hartman, M. Pan, and D. P. Lettenmaier, 2026: Model calibration and data assimilation for the National Water Model: A case study over California. Journal of Hydrometeorology, 27(1), 49–62. https://doi.org/10.1175/JHM-D-25-0066.1.

Ma, X., J. Wang, S. Gharari, N. Mizukami, M. Pan, and D. P. Lettenmaier, 2026: Monitoring reservoir storage using SWOT satellite observations and reservoir operation models. Water Resources Research, 62, e2025WR041223. https://doi.org/10.1029/2025WR041223.

Ma, X., D. Li, Y. Fang, S. A. Margulis, and D. P. Lettenmaier, 2023: Estimating spatiotemporally continuous snow water equivalent from intermittent satellite observations: An evaluation using synthetic data. Hydrology and Earth System Sciences, 27(1), 21–38. https://doi.org/10.5194/hess-27-21-2023.

Ban, Z., C. Xin, Y. Fang, X. Ma, D. Li, and D. P. Lettenmaier, 2023: Snowmelt-radiation feedback impact on Western U.S. streamflow. Geophysical Research Letters, 50(23). https://doi.org/10.1029/2023GL105118.

Li, D., R. A. Engel, X. Ma, E. Porse, J. D. Kaplan, S. A. Margulis, and D. P. Lettenmaier, 2021: Stay-at-home orders during the COVID-19 pandemic reduce urban water use. Environmental Science & Technology Letters, 8(5), 431–436. https://doi.org/10.1021/acs.estlett.0c00979.

SELECTED PRESENTATIONS

Ma, X., A. Liu, and D. P. Lettenmaier, 2025: Deep Learning and Reinforcement Learning for Reservoir Storage Forecasting Enhanced by SWOT: A Comparison with NWM–ESP. AGU Fall Meeting, New Orleans, U.S., Dec 2025.

Ma, X., and D. P. Lettenmaier, 2024: Monitoring Reservoir Storage Changes Using SWOT Observations and Land Surface Models. AGU Fall Meeting, Washington, D.C., U.S., Dec 2024.

Ma, X., and D. P. Lettenmaier, 2023: Do Trends in Global Precipitation Confirm Acceleration of the Global Hydrological Cycle? AGU Fall Meeting, San Francisco, U.S., Dec 2023.

Ma, X., and D. P. Lettenmaier, 2022: Reconstruction of Historical California Flood Forecasts. AGU Fall Meeting, Chicago, U.S., Dec 2022.

Ma, X., D. Li, and D. P. Lettenmaier, 2021: The Utility of Infrequent Along-Track Satellite Observations for Deriving Spatially Continuous Estimates of Snow Water Equivalent. AGU Fall Meeting, New Orleans, U.S., Dec 2021.

Xin, C., Y. Fang, D. Li, X. Ma, Z. Ban, H. Sun, S. J. Turnbull, and D. P. Lettenmaier, 2020: Quantifying the Impact of Declining Snow on River Discharge in a Warming Climate. AGU Fall Meeting, online, Dec 2020.

HONORS

Dissertation Year Award	University of California, Los Angeles	2025
Doctoral Student Travel Grant	University of California, Los Angeles	2023
Graduate Student Fellowship	University of California, Los Angeles	2021
Undergraduate Academic Merit Scholarship	Hohai University	2016–2018
National Innovation and Entrepreneurship Training Program, Research Fellowship	Hohai University	2016–2017
Undergraduate Outstanding Student	Hohai University	2017
Undergraduate Spiritual Scholarship	Hohai University	2017
Undergraduate Outstanding Student Leader	Hohai University	2016
Undergraduate Social Work Scholarship	Hohai University	2016

Chapter 1. Introduction

Hydrologic predictability depends fundamentally on two complementary capabilities: models that represent how water moves through the terrestrial system and observations that constrain the state of that system. Improvements in either capability can advance prediction, but their greatest value often emerges when they are used together. This dissertation is organized around that hypothesis. It consists of two studies. The first study (Chapter 2) examines advances in hydrologic modeling for short-range flood prediction; the second (Chapters 3 and 4) examines advances in satellite observation of reservoir storage change and their use to support seasonal reservoir forecasting and water-management decisions. These questions are especially important in the western United States (WUS), where strong seasonality, atmospheric-river storms, snow accumulation and melt, and recurrent droughts create rapid transitions between water surplus and scarcity (Dettinger and Anderson, 2015; Spinti et al., 2023).

Chapter 2 concerns the evolution of hydrologic models and their ability to improve flood prediction. Operational flood prediction in the United States developed from early conceptual rainfall-runoff models, and in particular the Stanford Watershed Model Crawford and Linsley, 1966, which represents interacting watershed processes including soil moisture, evapotranspiration, runoff generation, and channel response. The conceptual approach embedded in the Stanford Watershed Model helped establish the basis for continuous hydrologic simulation and later was the basis for operational river-forecasting systems, in particular the U.S. National Weather Service River Forecast System (NWSRFS). A major milestone was the Sacramento Soil Moisture Accounting model (SAC-SMA) described by Burnash et al. (1973), which was a derivative of the Stanford Watershed Model and became the core of the NWSRFS which has since (and continues to be) used extensively in operational flood forecasting in the U.S. and elsewhere.

Despite their relatively simple conceptual structure, well-calibrated operational models such as SAC-SMA remain difficult benchmarks to outperform. Their skill reflects not only model formulation, but also decades of calibration, local experience, and refinement of operational procedures (Smith et al., 2004, 2013; Moser et al., 2013). More recent spatially distributed land-surface and hydrologic models represent spatial variability in soils, vegetation, snow, and topography more explicitly, but greater physical and spatial complexity does not automatically guarantee better forecasts (Reed et al., 2004). Forecast skill still depends on meteorological forcing, model parameters, and estimates of the initial hydrologic state. Errors in precipitation forecasts affect flood magnitude and timing, while uncertain soil moisture and snow conditions influence how precipitation is converted to runoff (Adams and Dymond, 2019; Chifflard et al., 2018; Yin et al., 2022).

The National Water Model (NWM), developed by the National Oceanic and Atmospheric Administration and others, represents this newer generation of hydrologic prediction systems (Cosgrove et al., 2024). It uses the WRF-Hydro framework coupled with the Noah land surface model with multiparameterization options (Noah-MP) to provide spatially distributed hydrologic simulations and forecasts across the United States (NOAA, 2016; Niu et al., 2011; Gochis et al., 2020). This leads to the motivating question for my first study: can a modern spatially distributed model, when properly calibrated and supplied with high quality forcings and initial conditions, match or exceed the forecast accuracy of a mature operational system like NWSRFS? Archived forecasts from the California-Nevada River Forecast Center (CNRFC), which were produced using the long-established NWSRFS operational forecasting framework, provide a benchmark for answering this question (Su et al., 2025).

My second study concerns the parallel evolution of hydrologic observations. Reservoirs are central to water management because they regulate streamflow for water supply, flood control, hydropower, recreation, and environmental needs (Adam et al., 2007; Biemans et al., 2011; Hanazaki et al., 2022). Effective reservoir monitoring and operation require timely information about reservoir storage change, yet continuous in situ observations are unavailable for many reservoirs, especially in data-limited regions of the world (Gao et al., 2012). Satellite radar altimetry has provided an important alternative by measuring water-surface elevation (WSE) for inland water bodies, but conventional nadir altimeters sample only along narrow ground tracks, limiting their spatial coverage and revisit opportunities for lakes and reservoirs (Berry et al., 2005; Gao, 2015).

Recognition of these limitations motivated efforts to observe terrestrial surface water more comprehensively from space. Alsdorf et al. (2007) emphasized the need for global observations of spatial swaths (rather than tracks) of water-surface elevation (WSE), inundated area, and their spatial and temporal variations, and described wide-swath interferometric approaches that could overcome the sparse sampling of conventional altimetry. These developments ultimately led to the Surface Water and Ocean Topography (SWOT) mission (launched in late 2022), which provides spatially extensive measurements of both WSE and surface area (Biancamaria et al., 2016; Fu et al., 2024). For reservoirs, the combination of WSE and area creates new opportunities to estimate reservoir storage and storage change more consistently than with earlier satellite systems. The second study therefore asks whether this improved observing capability is sufficiently accurate and temporally useful for reservoir monitoring, particularly in data sparse regions. The first aspect of the second study (Chapter 3) therefore investigates the accuracy and reliability of SWOT-derived water surface level, surface area, and storage change estimates, and further examines whether a reservoir operation model constrained by SWOT observations can fill temporal gaps between satellite overpasses. Establishing whether these satellite-derived reservoir states are sufficiently accurate for monitoring provides the foundation for examining their value in reservoir prediction.

The second aspect of my work addressing the role and potential of improved hydrologic observations connects improved observations with improved models: specifically, improved observations are most useful for prediction when they can be incorporated into models. Monitoring current reservoir storage is an important first step (which I examine in Chapter 3), but water managers ultimately require forecasts of future storage and release (the topic of Chapter 4). Reservoir-operation models can propagate satellite-derived storage between SWOT overpasses, while inflow forecasts provide information about how the system may evolve beyond the most recent observation. Seasonal reservoir forecasts are therefore controlled by both the initial storage state and future inflow uncertainty. Initial-state errors tend to matter most at short lead times, whereas inflow uncertainty accumulates and can become increasingly important at longer lead times (Zhao et al., 2011; Yang et al., 2021). This model-observation integration is especially relevant in data-limited regions, where satellite observations and global forecast products may provide the only consistent information available for reservoir prediction.

Taken together, the two studies define the central argument of this dissertation: advances in hydrologic models improve the representation and prediction of system dynamics, while advances in Earth observation improve knowledge of the hydrologic state, and combining the two creates new opportunities for forecast-informed water management. The first study evaluates whether model calibration, precipitation bias correction, and discharge-informed adjustment of initial soil moisture improve short-range Noah-MP flood forecasts. The second assesses the accuracy of SWOT-based reservoir WSE, surface area, and storage change estimates, and integrates these observations into a reservoir storage forecast system that uses SWOT-derived reservoir storage to initialize seasonal reservoir-operation forecasts and evaluates predicted storage and release in terms of downstream agricultural water demand.

Chapter 2 evaluates Noah-MP flood forecasting in the Russian River, Yuba-Feather, and Santa Ana River basins in California, which span contrasting hydroclimatic conditions. Chapter 3 evaluates SWOT WSE, surface-area, and storage change estimates for 12 reservoirs in California, the Upper Colorado River Basin, and the Columbia River Basin where there exists high quality in situ observations for evaluation of SWOT reservoir products. Chapter 4 then extends the progression from observation to prediction by combining SWOT-derived initial storage, two seasonal streamflow forecast schemes, and a reservoir-operation model. Initially for two well-observed California reservoirs and then to the Tendaho Reservoir in Ethiopia, where observations are limited. Together, the two studies progress from improved hydrologic models, to improved hydrologic observations, to their integration for forecasting and management across both data-rich and data-limited environments.

References

Adam, J. C., Haddeland, I., Su, F., and Lettenmaier, D. P. (2007). Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' Rivers. *Journal of Geophysical Research*, **112**(D24).

Adams, T. E., III, and Dymond, R. L. (2019). Possible hydrologic forecasting improvements resulting from advancements in precipitation estimation and forecasting for a real-time flood forecast system in the Ohio River Valley, USA. *Journal of Hydrology*, **579**, 124138. https://doi.org/10.1016/j.jhydrol.2019.124138

Alsdorf, D. E., Rodríguez, E., and Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, **45**, RG2002. https://doi.org/10.1029/2006RG000197

Berry, P. A. M., Garlick, J. D., Freeman, J. A., and Mathers, E. L. (2005). Global inland water monitoring from multi-mission altimetry. *Geophysical Research Letters*, **32**(16). https://doi.org/10.1029/2005GL022814

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, **37**, 307–337. https://doi.org/10.1007/s10712-015-9346-y

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., et al. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, **47**(3). https://doi.org/10.1029/2009WR008929

Burnash, R. J. C., Ferral, R. L., and McGuire, R. A. (1973). *A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers*. U.S. Department of Commerce, National Weather Service.

Chifflard, P., Kranl, J., Zur Strassen, G., and Zepp, H. (2018). The significance of soil moisture in forecasting characteristics of flood events: A statistical analysis in two nested catchments. *Journal of Hydrology and Hydromechanics*, **66**(1), 1–11. https://doi.org/10.1515/johh-2017-0037

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., et al. (2024). *NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling*. **JAWRA Journal of the American Water Resources Association, 60**(2), 247–272. https://doi.org/10.1111/1752-1688.13184

Crawford, N. H., and Linsley, R. K. (1966). *Digital Simulation in Hydrology: Stanford Watershed Model IV*. Technical Report No. 39, Department of Civil Engineering, Stanford University.

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management, 111*(2), 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

Dettinger, M. D., and Anderson, M. L. (2015). Storage in California's reservoirs and snowpack in this time of drought. *San Francisco Estuary and Watershed Science*, **13**(2). https://doi.org/10.15447/sfews.2015v13iss2art1

Fu, L.-L., et al. (2024). The Surface Water and Ocean Topography Mission: A breakthrough in radar remote sensing of the ocean and land surface water. *Geophysical Research Letters*, **51**, e2023GL107652. https://doi.org/10.1029/2023GL107652.

Gao, H. (2015). Satellite remote sensing of large lakes and reservoirs: From elevation and area to storage. *WIREs Water*, **2**(2), 147–157. https://doi.org/10.1002/wat2.1065

Gao, H., Birkett, C., and Lettenmaier, D. P. (2012). Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, **48**(9). https://doi.org/10.1029/2012WR012063

Gochis, D. J., et al. (2020). *The WRF-Hydro Modeling System Technical Description* (Version 5.2.0). National Center for Atmospheric Research.

Hanasaki, N., Kanae, S., & Oki, T. (2006). *A reservoir operation scheme for global river routing*

models. Journal of Hydrology, 327(1–2), 22–41. DOI: 10.1016/j.jhydrol.2005.11.011.

Hanazaki, R., Yamazaki, D., and Yoshimura, K. (2022). Development of a reservoir flood control scheme for global flood models. *Journal of Advances in Modeling Earth Systems*, **14**(3), e2021MS002944. https://doi.org/10.1029/2021MS002944

Ma, X., Hartman, R. K., Pan, M., and Lettenmaier, D. P. (2026a). Model calibration and data assimilation for the National Water Model: A case study over California. *Journal of Hydrometeorology*, **27**, 49–62. https://doi.org/10.1175/JHM-D-25-0066.1

Ma, X., Wang, J., Gharari, S., Mizukami, N., and Lettenmaier, D. P. (2026b). Monitoring reservoir storage using SWOT satellite observations and a reservoir operation model. *Water Resources Research*, **62**, e2025WR041223. https://doi.org/10.1029/2025WR041223

Moser, C. L., Kroczynski, S., and Hlywiak, K. (2013). Comparison of the SAC-SMA and API-CONT hydrologic models at several Susquehanna River headwater basins. *Eastern Region Technical Attachment 2013-01*.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, **116**, D12109. https://doi.org/10.1029/2010JD015139

NOAA. (2016). *National Water Model: Improving NOAA's Water Prediction Service*. National Oceanic and Atmospheric Administration.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., & DMIP Participants. (2004). *Overall Distributed Model Intercomparison Project results*. Journal of Hydrology, 298(1–4), 27–60. https://doi.org/10.1016/j.jhydrol.2004.03.031

Smith, M., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., and Cong, S. (2004). The Distributed Model Intercomparison Project (DMIP): Motivation and experiment design. *Journal of Hydrology*, **298**, 4–26. https://doi.org/10.1016/j.jhydrol.2004.03.040

Smith, M., et al. (2013). The Distributed Model Intercomparison Project—Phase 2: Experiment design and summary results of the western basin experiments. *Journal of Hydrology*, **507**, 300–329. https://doi.org/10.1016/j.jhydrol.2013.08.040

Spinti, R. A., Condon, L. E., and Zhang, J. (2023). The evolution of dam-induced river fragmentation in the United States. *Nature Communications*, **14**, 3820. https://doi.org/10.1038/s41467-023-39194-x

Su, L., Pan, M., Hartman, R. K., and Lettenmaier, D. P. (2025). Improving National Water Model flood forecast skill over coastal western U.S. river basins. *Journal of Hydrometeorology*, **26**, 1129–1146. https://doi.org/10.1175/JHM-D-24-0072.1

Yin, D., Xue, Z. G., Bao, D., Rafieeinasab, A., Huang, Y., Morales, M., and Warner, J. C. (2022). Understanding the role of initial soil moisture and precipitation magnitude in flood forecast using a hydrometeorological modeling system. *Hydrological Processes*, **36**, e14710. https://doi.org/10.1002/hyp.14710

Zhao, T., et al. (2011). Effect of streamflow forecast uncertainty on real-time reservoir operation. *Advances in Water Resources*, **34**(4), 495–504. https://doi.org/10.1016/j.advwatres.2011.01.004

Chapter 2. Model Calibration and Data Assimilation for the National Water Model: A Case Study over California

This chapter has appeared in print in the *Journal of Hydrometeorology* in January 2026 as doi: 10.1175/JHM-D-25-0066.1.

Abstract

Flooding causes significant loss of life and property in the U.S. Improved flood forecasts have the potential to reduce these losses. NOAA and collaborators have developed the National Water Model (NWM), a nationwide hydrological prediction system intended for various applications, including flood forecasting. The NWM core is Noah-MP, a spatially distributed hydrological model that predicts river discharge given observed (or forecasted) precipitation, surface air temperature, and other surface meteorological variables. However, it has had only limited evaluation for flood forecasting in the western U.S. to date. Here, we report on retrospective flood forecasting experiments using the Noah-MP model, application of which we enhanced through model calibration, bias correction of precipitation, and adjustment of initial soil moisture through data assimilation of discharge observations. Our study focuses on selected flood events between 2004 and 2023 across three Forecast Informed Reservoir Operations (FIRO) watersheds in California, the Russian, Yuba-Feather, and Santa-Ana River Basins. Our results indicate that in the two humid and semi-humid basins, reforecasts with Noah-MP produced flood forecast results with accuracy comparable to archived forecasts produced by the National Weather Service California-Nevada River Forecast Center (CNRFC). With precipitation bias correction and adjustments to initial soil moisture, Noah-MP reforecast results

were slightly more accurate than CNRFC forecasts in all three river basins, particularly for lead times exceeding two days.

1. Introduction

Flooding is among the most dangerous and costliest of natural disasters in the U.S. (Center for Research on the Epidemiology of Disasters, 2017). Damage and loss of life occurs from events ranging in magnitude from flash floods to flooding of the largest rivers, and affects both rural and urban areas (Hammond et al. 2015; Sene 2016). The NOAA (National Oceanic and Atmospheric Administration) data for Billion-Dollar Weather and Climate Disasters over the last 44 years indicates that flooding costs in the US average about 4.5 $B annually. Some recent work however (U.S. Joint Economic Committee, 2024) suggests that these estimates are substantially too low.

Data from the National Weather Service (NWS), covering both the 10-year period from 2009 to 2018 and the 30-year period from 1989 to 2018, indicate that floods rank second in weather-related fatalities across the United States, trailing only heat waves (National Weather Service, 2023). Of concern is that the frequency of 'billion-dollar disasters' and the total cost of flood-related damages has been rising over time (Neri et al. 2020; Slater and Villarini 2016). Recently, floods have become less deadly, presumably due to better and longer lead warnings, especially for flash floods (Jonkman et al. 2024). More accurate and robust flood forecasts are central to providing earlier and more precise information, that can help to reduce flood damage and loss of life (Adams 2016; Jain et al. 2018; Yatheendradas et al. 2008).

Flood forecasts in California are issued by the California-Nevada River Forecast Center (CNRFC), which is an arm of NOAA's National Weather Service. CNRFC produces Quantitative Precipitation Estimates (QPE) as well as Quantitative Precipitation Forecasts (QPF) every 6 hours, which are an integral part of its flood forecast operations. QPFs currently have lead times of 6 hours to 10 days. CNRFC uses QPEs and QPFs to force a snow accumulation and ablation model (SNOW-17; Eric Anderson, 2006) and then the rainfall-runoff model that originated as the so-called Sacramento model (Burnash et al., 1973) with various updates over time (Raleigh and Lundquist 2012; Zhang et al. 2012a); the two models are the core of the CNRFC (and NWS) flood forecasting system. The hydrological models, together with infrastructure that handles the required parameters, states, data, and other hydrologic operations (e.g. routing, reservoir operations, etc.) form the CNRFC's Community Hydrologic Prediction System (CHPS).

Here, we investigate several alternatives to this system which we evaluate via comparisons of reconstructed flood forecast accuracy with the accuracy of historical flood forecasts archived by CNRFC. The first is the rainfall-runoff model around which the CNRFC system is constructed. As noted above, the Sacramento model (also known as the Sacramento Soil Moisture Accounting model, or SAC-SMA) was developed many years ago. Although effective for flood forecasting and other applications (Chouaib et al. 2021; Alvarez and Barco 2023; Bahrami et al. 2023; Wi and Steinschneider 2022), it doesn't represent the effects of vegetation on evapotranspiration and precipitation interception directly, doesn't physically link moisture exchanges between soil layers, and simplifies key processes like evapotranspiration (Agnihotri and Coulibaly 2020; Zhang et al. 2012b). Furthermore, it is spatially lumped, hence is limited in representation of spatial variations in surface and subsurface hydrologic characteristics. These limitations potentially reduce flood forecasting accuracy on a catchment-scale (Uliana et al. 2019; Roodsari et al. 2019). Furthermore, CNRFC at present doesn't incorporate a formal data assimilation system into its forecasts, although an informal adjustment procedure is used to manually adjust forecasts based on a priori streamflow "nowcast" errors at the time of forecast. To the extent that CNRFC describes operational flood forecasting as "the art of balancing experience, current conditions analysis, and model guidance" (California Nevada River Forecast Center, n.d.) reliance on manual processes may constrain consistency and accuracy in forecasts (Eckhardt et al. 2005; Ndiritu 2009).

Notwithstanding these potential limitations, SAC-SMA has generally been competitive with more modern models, particularly when well-calibrated (Sorooshian et al. 1993). Intercomparison studies such as the Distributed Model Intercomparison Project (DMIP) (Smith et al. 2004, 2013) have demonstrated that lumped models like SAC-SMA can perform comparably to or even outperform distributed models, especially when forecasts are primarily focused on basin outlets and when high-resolution input data required by distributed models are not readily available (Smith et al. 2004, 2013). There have, however, been arguments to the effect that more physically-based and spatially distributed models should have the potential to improve flood forecast accuracy (Moser et al.; Daliakopoulos and Tsanis 2016). Nonetheless, this potential has yet to be demonstrated in the context of operational forecasts.

The National Water Model (NWM) (NOAA, 2016) is intended to provide a comprehensive framework for hydrologic forecasts on a national scale. It is based on the WRF-Hydro system, which combines representations of land-atmosphere interactions as represented by the Weather Research and Forecasting (WRF) mesoscale atmospheric model (Gochis et al., 2020), coupled with hydrological processes represented by Noah-MP (Noah with multi-parameterization) (Niu et al. 2011) as its hydrologic core. Noah-MP is a spatially distributed land surface (hydrology) model that allows users to choose multiple options for various physical processes (Niu et al. 2011; Cai et al. 2014; Salamanca et al. 2018; Zhong et al. 2021; Zhuo et al. 2019). Noah-MP supports more detailed parameterizations of key hydrologic processes like runoff and soil moisture (as compared with LSMs that focus more exclusively on land-atmosphere energy and moisture exchanges). Its complexity and flexibility suggests its potential as an alternative to SAC-SMA, particularly in areas where the hydrological processes are complex such as in the topographically diverse western U.S. (Zhang et al. 2012; Bournas and Baltas 2021), although this has yet to be demonstrated in an operational context.

In a hydrological forecast setting, the WRF-related fluxes to Noah-MP are usually replaced by uncoupled forcings, notably of precipitation (which comes from QPE and QPF). Hence, while WRF-Hydro and NWM have been used for applications of Noah-MP in an off-line setting, here we refer the off-line implementation of NWM (which we use here) simply as Noah-MP. This is the structure that was used by Su et al. (2024) to compare Noah-MP-based flood forecasts in a reforecast setting with RFC (CNRFC and Northwest RFC) archived forecasts along the coastal Western U.S. They found that Noah-MP has comparable forecasting performance to RFC reforecasts in the northern basins but performed worse in the southern basins. Both models underestimated flood peaks, with Noah-MP showing greater discrepancies at longer lead times (> 48 hours) and a tendency to predict earlier peak timings. In addition, they suggested that a plausible contributor to forecast errors was inaccuracies in the predicted precipitation (i.e., QPF). Also, antecedent initial soil moisture conditions may play an equally important role as precipitation in flood forecasting, especially in relatively dry basins (Chifflard et al. 2017; Yin et al. 2022). To our knowledge, more comprehensive research on Noah-MP-based flood forecasting is still lacking, particularly in terms of performance improvements through model parameter estimation, input forcings, and initial conditions under various climate conditions.

Accordingly, we focus here on potential improvements in the flood forecast accuracy of Noah-MP relative to SAC-SMA in California watersheds through a) model calibration, b) mitigating QPF biases, and c) improved initialization of soil moisture through data assimilation. In the remainder of this paper, we report results of a comprehensive analysis of the accuracy of Noah-MP reforecasts conducted in a framework as similar to that used in operational forecasting as possible, relative to historical operational forecasts archived by CNRFC that were produced using CHPS in real-time. We focus in particular on the three above-mentioned FIRO watersheds, which span the range of hydroclimatic conditions encountered over California: a minimally snow-affected semi-humid northern coastal watershed (Russian River), a mixed rain-snow northern California mountainous watershed (Yuba-Feather) and a semi-arid southern California watershed (Santa Ana). Using a calibrated version of Noah-MP, we adjust precipitation inputs and initial soil conditions to enhance forecast accuracy. Section 2 describes the study domains and methods, while Section 3 details the experimental design. Section 4 evaluates both flood reconstructions (simulations) and reforecasts. Finally, Section 5 discusses the findings and potential future directions, with concluding remarks presented in Section 6.

2. Study description and methods

2.1 Study area and forecast points

Our study area consisted of three FIRO (forecast-informed reservoir operations) river basins in California. FIRO is an evolving reservoir-operations strategy that is intended to improve reservoir management by integrating next generation weather (especially precipitation) and hydrologic forecasts into reservoir operating policies (American Meteorological Society; 2020). Within California, three FIRO watersheds have been identified for collaborative studies among California's Department of Water Resources, the National Weather Service, the U.S. Army Corps of Engineers, and local water agencies and stakeholders: the Russian, the Yuba-Feather, and the Santa-Ana. The three watersheds experience a range of climate conditions. The Russian River basin in Northern California is a coastal watershed with a drainage area of about 3850 km^2. Winters are generally wet and summers dry, with most precipitation occurring from October to April. Although there are two reservoirs in the basin (Lake Mendocino and Lake Sonoma), their impact on basin outflows is minimal compared to flood magnitudes (Su et al., 2024). The Yuba and Feather Rivers in Northern California (combined drainage area 13770 km^2) receive most of their annual rainfall from November to April, largely from atmospheric river storms that bring heavy rain and snowfall above about 1500 m elevation (Ralph et al. 2022). The Santa Ana River Basin is the largest watershed in Southern California, with a drainage area of over 6784 km^2 in San Bernardino, Riverside, Orange, and parts of Los Angeles Counties (CLARKE 1996; Miller 2020). It is characterized by coastal mountains and dry alluvial valleys. Average annual precipitation is slightly under 400 mm with most rainfall occurring between November and March, leading to generally small river flows except during the winter season, when intense rainfall can trigger flash floods (Sene 2016b). With a population of approximately 4.8 million, primarily in urban areas like Riverside and Santa Ana, accurate flood forecasting is critical both to avoiding flood-related property damage and loss of life.

We focused on six CNRFC forecast points in the three FIRO regions, three in the Russian River basin, two in the Yuba Feather, and one in the Santa Ana (see Figure 1 and Table 1). All of the forecast points are only modestly influenced by human activities (in the case of the Santa Ana, the forecast point is upstream of major population centers). However, it should be noted that the Russian River at Hopland is significantly regulated by Lake Mendocino, which controls approximately one-third of its drainage area and is specifically operated to reduce flood impacts at Hopland. Together, Lake Mendocino and Lake Sonoma regulate around 15% of the Russian River drainage area above Guerneville. Despite this regulation, its impact on the current analysis might be limited, as both reservoirs typically reduce releases to minimum discharge levels during significant runoff events. Aside from the upstream-most portions of the Yuba-Feather, the three basins are rain-dominated. We selected the forecasting points based on the availability of sub-daily streamflow records and archived forecasts provided by the CNRFC, which has the responsibility for flood forecasting in the three FIRO watersheds. Table 1 gives the drainage areas, stream gauge identifiers, and CNRFC forecast point designators for watersheds and forecast points. Note that only the locations in the Russian River Basin are official NWS flood forecast points that operationally receive additional attention and scrutiny. The locations in the Yuba are intended to supplement downstream forecasts for the Yuba at Marysville. The forecast point in the Santa Ana is intended to support flash flood guidance issued by NWS Weather Forecast Offices.

Table 1 Drainage area, observing sites, and RFC sites for the study region.

Basin	Site Name	Drainage Area [km^2]	OBS Sites	RFC Sites	CNRFC Forecasts Periods
Russian River Basin	Guerneville	3425	11467000	GUEC1	Dec 2002 - Feb 2019
	Healdsburg	2030	11464000	HEAC1	Dec 2002 - Feb 2019
	Hopland	926	11462500	HOPC1	Dec 2002 - Feb 2019
Yuba-Feather River Basin	Smartsville	217	11418500	DCSC1	Dec 2014 - Apr 2020
	Middle Yuba River Below Our House Dam	371	CDEC ORH	OURC1	Jan 2016 - Apr 2019
Santa Ana River Basin	Lytle C A Colton	476	11065000	YTLC1	Dec 2014 - Feb 2019

2.2 Land Surface Model and Data sets

2.2.1 Hydrology model

We used the Noah-MP hydrology/land surface model within the WRF-Hydro framework (version 5.2.0). We implemented the model in an uncoupled configuration, meaning that atmospheric forcings were externally supplied (effectively bypassing the role of the WRF land-atmosphere model). We used a spatial resolution of 6 km in the Russian and Yuba-Feather basins, and a finer spatial resolution of 3 km in the Santa Ana River basin.

The primary prognostic variables in Noah-MP are surface runoff, soil water storage and drainage, evapotranspiration, snow processes (including snow water storage and melt), and aquifer recharge. Surface and subsurface flows are laterally routed through modules designed to account for gravitational effects driven by elevation, influencing both surface runoff and subsurface drainage before these flows are directed into the channel network, as described by Gochis et al. (2020). We utilized runoff option 3 (the default free drainage option, as per Schaake et al. (1996), which was evaluated in Su et al. (2024) and found to surpass other runoff options for flood forecasting performance under default Noah-MP settings and parameters.

2.2.2 Meteorological forcings

Noah-MP requires spatially distributed meteorological inputs, including precipitation, surface air temperature, specific humidity, surface pressure, wind speed, and both downward longwave and shortwave radiation. Among these, precipitation stands out as the most critical factor for accurate flood forecasting (Adams and Dymond 2019; Chiang et al. 2007; Sun et al. 2020).

We used the same precipitation forcing as do the RFCs so that we could better isolate the forecast skill of Noah-MP in comparison with RFC methods. Specifically, the precipitation we used for our baseline simulations, as well as our reforecasts, was QPE and QPF from CNRFC. The available time period for QPE and QPF is Jan 2001-Current. The lead time for CNRFC QPF is 6-72 hours prior to Sep 2012 and 144 hours thereafter. The initialization interval for CNRFC is once per day (12:00) before Oct 2010 and twice per day (12:00 and 18:00) after Oct 2010.

We interpolated the 4-km and 6-hourly QPE and QPF data to 6-km and 3-hourly resolutions. Forcings other than precipitation came from Pan et al. (2023). The Pan et al. (2023) data set provides 1-km, hourly, near real-time data for the contiguous United States, derived from the North American Land Data Assimilation System (NLDAS; Xia et al. 2012) and rescaled using the Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008). We aggregated the 1-km hourly data to 6-km and 3-hourly resolutions. Before running the baseline simulations, we performed a 2-year spin-up simulation using Pan et al. (2023) forcings, including precipitation, to minimize initialization effects, particularly in soil moisture and SWE.

3.0 Experimental design

We conducted a series of experiments to improve flood forecasting performance of Noah-MP, focusing on calibrated parameters (section 3.1), flood reforecasts (section 3.2), QPF bias correction (section 3.3), and data assimilation (section 3.5). First, we estimated static parameters for each forecast point through model calibration, contrasting these calibrated parameters with off-the-shelf settings to assess their impact. Using QPE for calibration, we then replaced QPE with QPF to perform reforecasts for historical CNRFC flood events. To address systematic biases in QPF, we applied bias correction, and to further improve forecasts, we assimilated discharge observations prior to flood peaks to adjust initial conditions.

3.1 Model parameter estimation

Following Su et al. (2024) we first chose the Noah-MP runoff parameterization option that resulted in the greatest increase in KGE. Consistent with Su et al. (2024) we found that option 3 (free drainage) was always the best. We then adjusted the spatially variable Noah-MP parameters using KGE as the objective function, for five parameters: saturated hydraulic conductivity (DKSAT), saturated soil moisture (SMCMAX or porosity), the pore-size distribution index (BEXP), the deep drainage coefficient (SLOPE), and the surface runoff parameter (REFKDT). We chose these parameters based on previous studies (Holtzman et al. 2020; Sun et al. 2020; Sofokleous et al. 2023; Quenum et al. 2022; Mascaro et al. 2023; Bass et al. 2023). While default settings for DKSAT, SMCMAX, and BEXP in Noah-MP are based on State Soil Geographic Database (STATSGO) soil types (NCAR, 2022), these parameters are subject to significant site-by-site uncertainty. Su et al. (2024) summarized these most sensitive parameters and their realistic limits during parameter estimation following Cai et al. (2014). We constrained adjustments to these parameters to be within realistic ranges (see Supplement for details).

We selected the flood events for which parameter estimation was targeted using peaks-over-threshold (POT). We used POT3, which means for each forecasting point, the floods we incorporated in the parameter objective function were the 3rd most extreme events in the observation record, where n is the length of the calibration period in years.

After identifying the POT flood events (see Table A1), in each river discharge record, we used the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker 2007) to calibrate the Noah-MP static parameters noted above. Su et al. (2024) found, in their application of DDS to Noah-MP, that DDS usually showed rapid improvement in an objective function similar to the one we used within the first 50-200 iterations, with diminishing returns beyond that. They therefore terminated their search after 250 iterations, a criterion that we also used. After each parameter adjustment iteration, the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) was used as the objective function to assess improvements compared to the previous iteration. We evaluated 6-hour forecasts against observations and calculated KGE values over the full flood periods, defined as 36 hours before and after flood peaks (72 hours in total) in the Russian and Yuba-Feather River Basins, and 18 hours before and after flood peaks (36 hours in total) in the Santa Ana River Basin. A KGE value of 1.0 represents perfect match between simulated results and observations, and a KGE value exceeding -0.41 indicates that the model performs better than simply forecasting the mean flow (Knoben et al. 2019). Note that the CNRFC recalibrates its model parameters roughly every six years (e.g., four times in the past 20 years for the Russian basin). Thus, archived CNRFC forecasts reflect parameters and data systems at the forecast issuance time, unlike our Noah-MP reforecasts that utilize current modeling configurations.

3.2 Historical flood reconstruction

After calibration, we ran Noah-MP with calibrated parameters and QPE forcings and archived the model states (soil moisture and SWE) at each time step for all flood events as the "warm" starting states. For each flood event, we initiated simulations from this "warm state" at the starting point of forecasts (from which point on we forced the model with QPF, with lead times of 1, 2, 3, and 4 days relative to the (observed) flood peak. In our QPF-based flood reconstruction runs, we kept all other forcings, including air temperature, specific humidity, surface pressure, wind speed, and both downward longwave and shortwave radiation, the same as in the calibration runs. For future applications, all other forcings can come from forecasts. Since precipitation is the dominant variable in flood forecasting, the use of either historical or forecasted data for other forcings is unlikely to result in significant differences. In the Yuba River Basin, there are likely some events where the forecasts are quite sensitive to errors in the air temperature (and thus rain-snow elevation) forecasts (Sumargo et al. 2020).

3.3 Bias correction of QPF

We compared the Noah-MP based flood reconstruction (section 3.2; based on original QPF) with CNRFC flood archives. Flood events were selected based on CNRFC operational flood forecasts, ensuring that both hits and false alarms were included in the analysis and avoiding selection bias associated with observation-based event selection. After that, we noticed that as lead times become longer, QPF tends to have a systematic downward bias in the forecasted peak precipitation relative to observations (for which we used QPE as a surrogate; see Figure 2). Consequently, the forecasted flood peaks and volumes tend to underestimate observations. To alleviate this problem, we used the historical QPFs and QPEs to rescale the storm hyetographs. We used half of the flood events to calculate the scaling magnitude for each time in the precipitation hydrograph, and the other half for validation. The validated scale factors were stored for future use in precipitation adjustment. For the "training" precipitation, we randomly selected one or two flood event each year such that the total number of events for training was about 50% of the total number of flood events. We first average all the training QPF hyetographs and the corresponding observed QPE hyetographs. Given the differences between the averaged QPE and QPF hyetographs, we scaled the average QPF value at each time step. We saved the scale factors and validated using the validation part of the QPF series. For the Russian and Yuba-Feather River Basins, the adjustment period was 36h before to 36h after the QPE peak time (total 72h); for Santa-Ana, a shorter period from 18h before to 18h after the QPE peak time (total 36h). While this approach assumes knowledge of the observed peak time—which is, of course, unavailable during forecasting—future forecasting applications can address this limitation by applying the same scale factors to QPF peaks, thereby adjusting QPF magnitudes accordingly.

3.4 Adjustment of antecedent soil moisture through data assimilation

Note that after bias-correcting QPF, we attribute forecasting errors primarily to uncertainty in initial soil moisture conditions. To further improve forecast accuracy, we used a particle filtering approach (Djuric et al. 2003) to optimize the initial soil moisture content at the first (top) Noah-MP soil layer (smc1) 24h before the forecasted peak discharge, denoted as t = -24h. This method aimed to adjust the initial smc1 iteratively based on observed discharge data within a short period before the forecasted flood peaks (e.g., from t = -24h to t = -12h). It should be noted that our method depends on identifying flood peaks occurring relatively early within the forecast period (24 hours). In practice, we estimate flood peak timing based on the original flood forecasts, and then use the 24-hour window preceding each forecasted peak to adjust initial model states. Consequently, this approach inherently relies on forecast-derived peak timing estimates rather than known peak times, and its applicability may decrease when the real flood peaks way different from the forecasts.

The process began with the generation of an ensemble of potential initial soil moisture states, known as particles. Here, we used 100 particles. Each particle represents a plausible initial condition sampled from a normal distribution centered around the original baseline smc1 value and is intended to capture the initial uncertainty. To enhance the efficiency of the assimilation process, the initial soil moisture values (smc1) were adjusted based on the observed bias. Specifically, if the simulated discharge was smaller than the observed value, the baseline smc1 was scaled up by a factor of 1.5; conversely, if the simulated discharge was larger, the baseline smc1 was scaled down by dividing by 1.5. The scaling factor of 1.5 was determined empirically through preliminary experiments, where it was found to consistently improve the alignment of simulated discharge with observations. This targeted adjustment allowed the initial ensemble of particles to better align with observed conditions while maintaining diversity through added random perturbations. The ensemble was designed to represent the range of potential initial conditions that could influence model performance during the simulation window.

For each particle, we conducted forward simulations over the specified window from t = -24h to t = -12h (for the Santa-Ana River Basin, the window is from t = -12h to t = -6h due to flash floods). We then compared model-simulated discharge to observed discharge values at corresponding time steps. Assimilation was only performed for flood events where the absolute error between simulated and observed discharge exceeded 50%, ensuring that adjustments were focused on significant discrepancies. To quantify the agreement between simulated and observed discharge, we computed weights for each particle using a Gaussian likelihood function. This function accounted for the differences between simulated and observed values, scaled by an assumed observational error variance (Crisan 2001).

To address particle degeneracy (Doucet et al. 2001) and maintain diversity, weights were normalized, and resampling was conducted to prioritize particles with higher weights that better matched observed discharge data. Rather than directly using the resampled particles, we derived final results by averaging the model output discharge for the 10 particles with the highest weights. This approach provided a refined initial condition for future simulations and ensured that the model's starting state was more closely aligned with observations. If the updated smc1 did not sufficiently improve the match between simulated and observed discharge, the process was repeated iteratively. This iterative process continued until the KGE value between simulated and observed flow during the period 24 to 12 hours (for the Santa-Ana River Basin, the period is 12 to 6 hours) before forecasted flood peak was larger than 0.7 or was terminated if no improvement in KGE occurred. If it never reached this threshold, we iterated 50 times and chose the smc1 that corresponded to the highest KGE value.

The data assimilation approach allowed for informed and automatic adjustments to the initial soil moisture condition by integrating information from observations prior to the forecast time. By iteratively refining the initial smc1, this approach provided a way to enhance the model's ability to simulate discharge accurately, ultimately improving the reliability of hydrological forecasting.

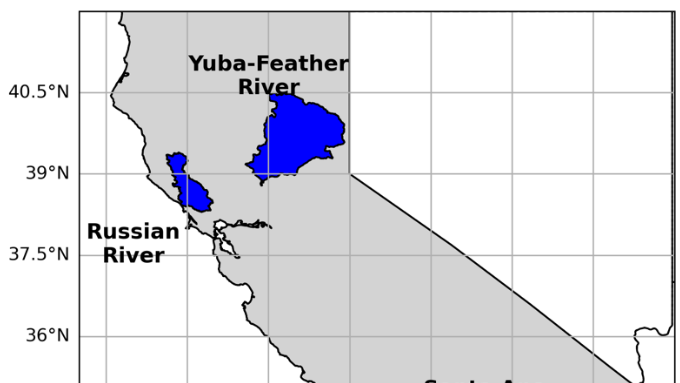


Figure 1. Map of study region showing the three FIRO watersheds.

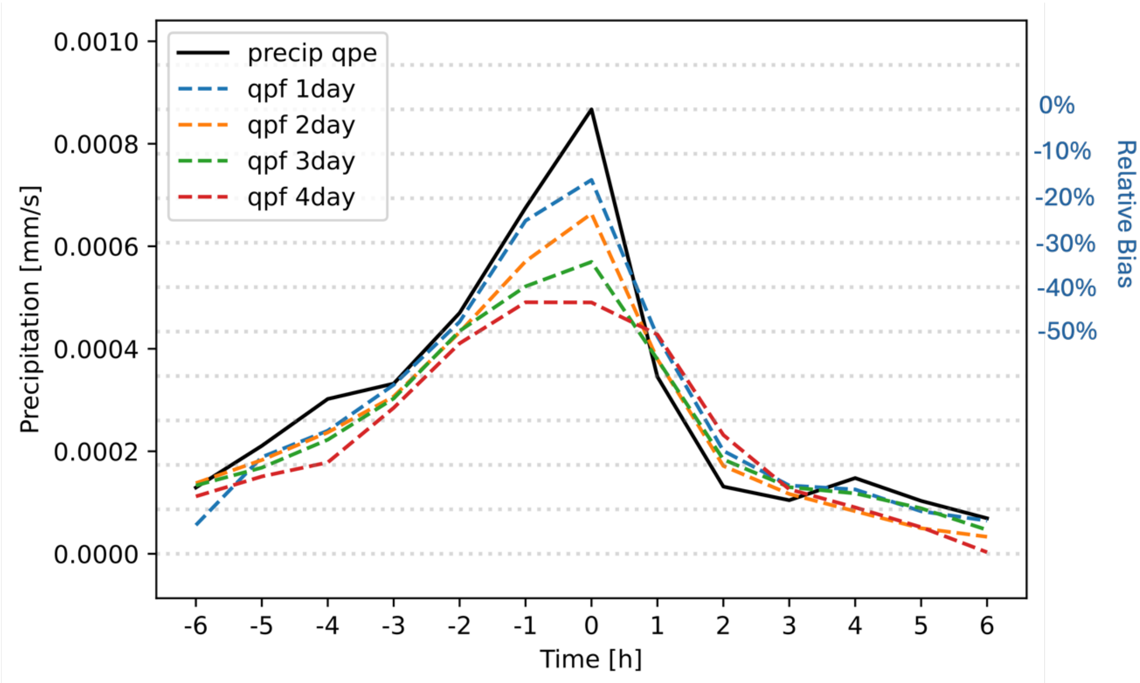


Figure 2. Normalized precipitation hydrograph weighted by QPE peaks, using Yuba-Feather DCSC1 as an example to illustrate the systematic bias in QPF relative to QPE.

4 Results

4.1 Streamflow simulation after model calibration

Scatterplots of POT3 flood forecasts produced by calibrated Noah-MP simulations for all six forecast points in the three FIRO watersheds are shown in Figure 3. The initial Noah-MP parameters were derived from the 30-arc-s hybrid STATSGO soil texture datasets (NCAR, 2022). After calibration, the KGE values for GUEC1, HEAC1, HOPC1, DCSC1, OURC1, and YTLC1 were 0.84, 0.87, 0.75, 0.83, 0.70, and 0.69, respectively. Calibration generally improved Noah-MP's streamflow simulation skill by 30% to 110% relative to simulations with "off the shelf" parameters. The performance of our flood flow simulations is notably strong when compared to past studies. For instance, Mascaro et al. (2023) documented an NSE of 0.59

(corresponds to KGE values around 0.70 – 0.75) in Noah-MP simulations relative to observed streamflow from 2008 to 2011. Similarly, Lin et al. (2018) assessed the streamflow simulation capabilities of Noah-MP using 271 USGS gauges across Texas, where their highest reported NSE was approximately 0.7 (the corresponding KGE is typically around 0.8 to 0.85). Our calibration results are comparable to Su et al. (2024) who employed the same DDS calibration algorithm that we used. Figure A1 shows the weighted average flood hydrographs for the six sites, with weights derived from the peak flood values. The results indicate no substantial biases in the simulated flood shapes, suggesting a good alignment between the observed and modeled hydrograph profiles.

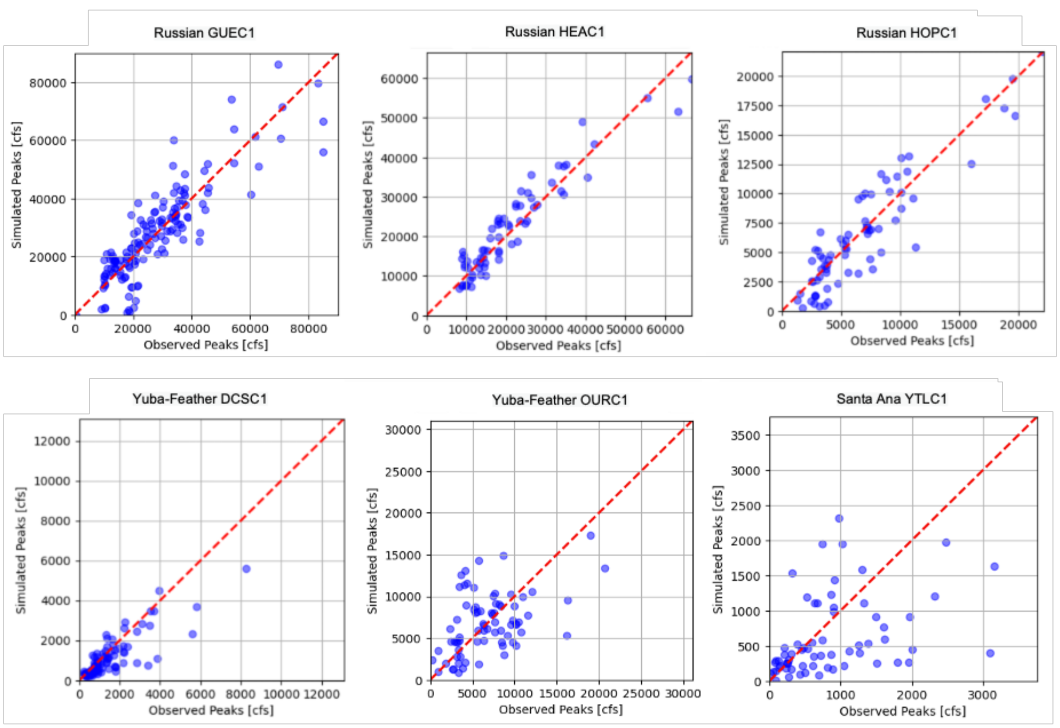


Figure 3. Scatter plot of simulated flood streamflow and observed flood streamflow in six forecasting points in the FIRO watersheds.

4.2 Flood reforecasts after QPF and soil moisture adjustments

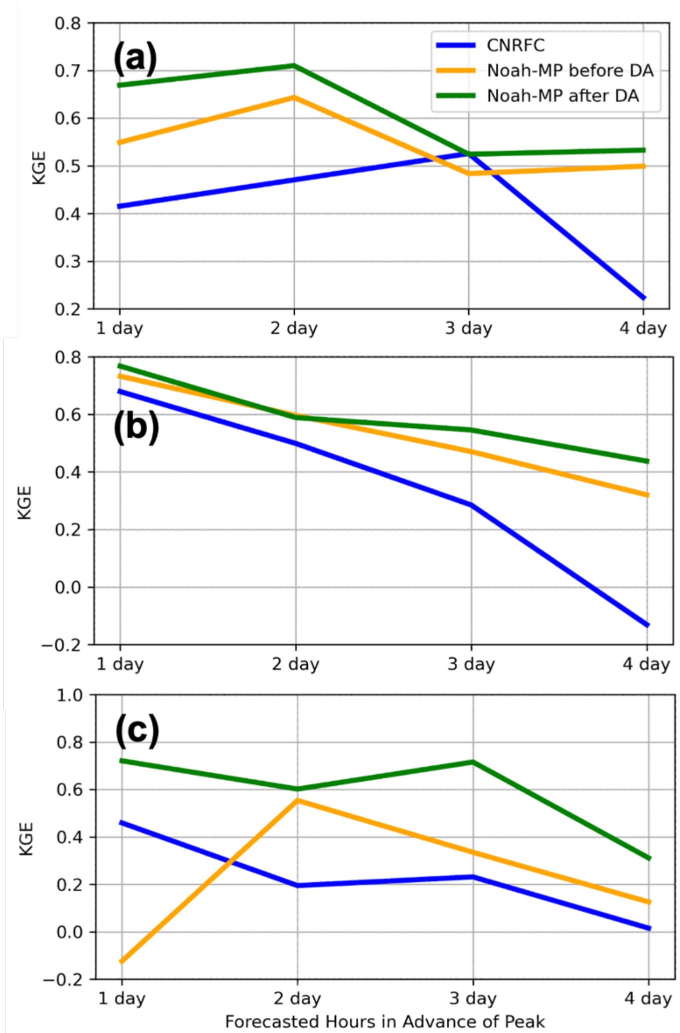


Figure 4. KGE for Russian (a), Yuba-Feather (b), and Santa-Ana (c) Rivers for lead times 1, 2, 3, and 4 days. CNRFC forecast results are shown in blue; Noah-MP reforecasts in yellow line; and Noah-MP reforecasts after QPF bias correction and data assimilation are shown in green.

Figure 4 shows the Kling-Gupta Efficiency (KGE) values for the three flood forecast points in the Russian (GUEC1), Yuba-Feather (DCSC1), and Santa Ana (YTLC1) River basins, for lead times 1, 2, 3, and 4 days before the observed peaks. Since the initial soil moisture adjustments through data assimilation were performed based on the results after QPF corrections, we collectively refer to the QPF correction and data assimilation processes as “Noah-MP after DA”. We computed the KGE values for all flood events during the full flood periods, defined as 72 hours around the peaks in the Russian and Yuba-Feather River Basins, and 36 hours around the peaks in the Santa Ana River Basin. The figure includes comparisons of CNRFC forecasts, Noah-MP forecasts after model calibration, and Noah-MP forecasts incorporating both QPF bias correction and data assimilation to improve initial soil moisture conditions.

Figure 4 indicate that forecast accuracy generally decreases with lead time beyond two days for all models and configurations (due to reduced QPF accuracy) Russian River CNRFC forecasts exhibit relatively strong performance at shorter lead times, but their performance deteriorates sharply beyond 2 days lead. In contrast, Noah-MP forecasts without data assimilation (DA) gradually improves up to 2 days lead time but their accuracy declines for longer leads. The inclusion of QPF bias corrections and initial condition adjustments through data assimilation reduces the rate of KGE decline beyond 3-day lead. Data assimilation and QPF bias correction not only enhance predictive accuracy but also effectively mitigate the sharp decline observed in both the CNRFC and unadjusted Noah-MP forecasts at longer lead times.

The Yuba-Feather basin results (Figure 4 panel b)) further underscore the advantages of incorporating QPF bias correction and initial soil moisture state adjustments. Noah-MP forecasts with these adjustments (Noah-MP after DA) consistently outperformed both uncorrected Noah-MP (Noah-MP before DA) and CNRFC forecasts for all lead times, with higher KGE values, particularly at 1- to 3-day forecast lead times. In contrast, the CNRFC forecast KGE values show sharp declines in KGE beyond 2-day lead times and negative KGE values for lead times beyond 4-days.

For the Santa Ana River basin, the efficacy of the adjustment of QPF bias correction and initial soil moisture condition updating becomes even more evident than for the two northern basins. At all lead times, the KGE values of Noah-MP after DA results are around 50% higher than those of the CNRFC forecasts (corresponding percentages are 5% to 20% in the two northern basins), meaning that the impacts QPF unbiasing and pre-flood soil moisture adjustments are more positive in this semi-arid basin. Overall, Noah-MP after DA outperforms both the uncorrected Noah-MP and CNRFC forecasts, particularly at 3- and 4-day lead times.

Overall, our results underscore the value of integrating QPF bias correction and initial condition updating in flood forecasting. The effect of these adjustments on forecast accuracy are especially evident at long forecast lead times. With these adjustments, Noah-MP based forecasts can outperform CNRFC flood peak magnitude forecasts.

(a) Peak Magnitude differences

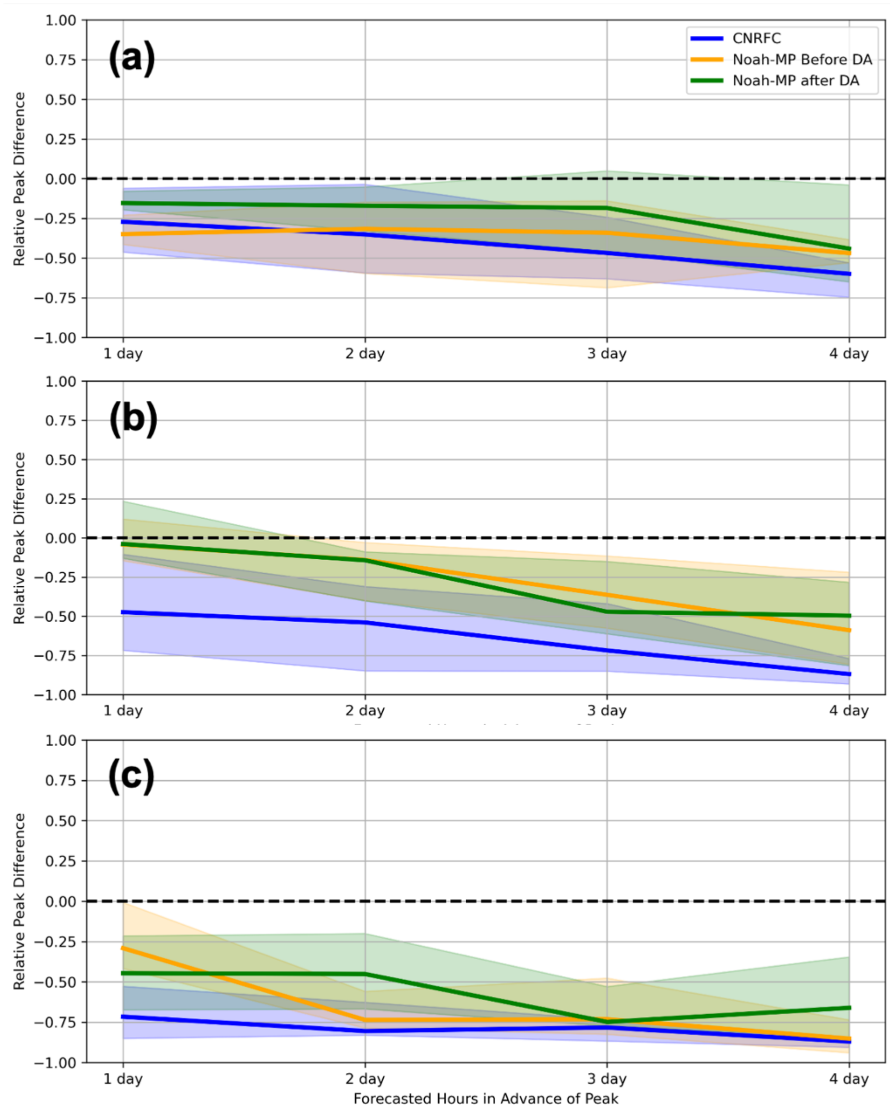


Figure 5. Median (solid lines) and interquartile range (shading) of the relative differences of floods peak (model minus observed divided by observed) streamflow of Noah-MP reforecasts and RFC forecasts as a function of forecast hours in advance of the observed peak (lead time) in Russian (a), Yuba-Feather (b), and Santa-Ana (c) for lead times for 1, 2, 3, and 4 days.

The previous section evaluated impacts on forecast accuracy using KGE taken over the period (36-h before and after flood peaks in the Russian (GUEC1) and Yuba-Feather (DCSC1) Basins; 18-h before and after flood peaks in the Santa-Ana (YTLC1) Basin). Here, we evaluate forecast accuracy based only on differences between observed and forecasted peak flow magnitudes. Figure 5 shows the medians and the interquartile range of relative peak differences, defined as the percentage difference between forecasted and observed peak flows, for all three rivers at 1- to 4-day lead times for CNRFC archived forecasts (blue), Noah-MP forecasts based on calibrated parameters(yellow), and Noah-MP after QPF corrections and initial condition adjustments (green). Overall, the enhanced Noah-MP forecasts show clear improvements in accuracy and reduced variability compared to both uncorrected Noah-MP and CNRFC results.

In all three river basins, CNRFC forecasts have a consistently negative bias, that becomes more severe at longer lead times. This likely is due to the effect of bias in the QPFs. The uncorrected Noah-MP forecasts also tend to underestimate peak flows but with slightly less bias than the CNRFC forecasts. Noah-MP accuracy for peak flow forecasts is improved with QPF and initial soil moisture adjustments (green). Although negative biases in Noah-MP forecasts still exist post-adjustment, the median values are closer to zero, and the variability (IQD) is substantially reduced. This trend is consistent across the Russian, Yuba-Feather, and Santa Ana basins.

(b) Peak time differences

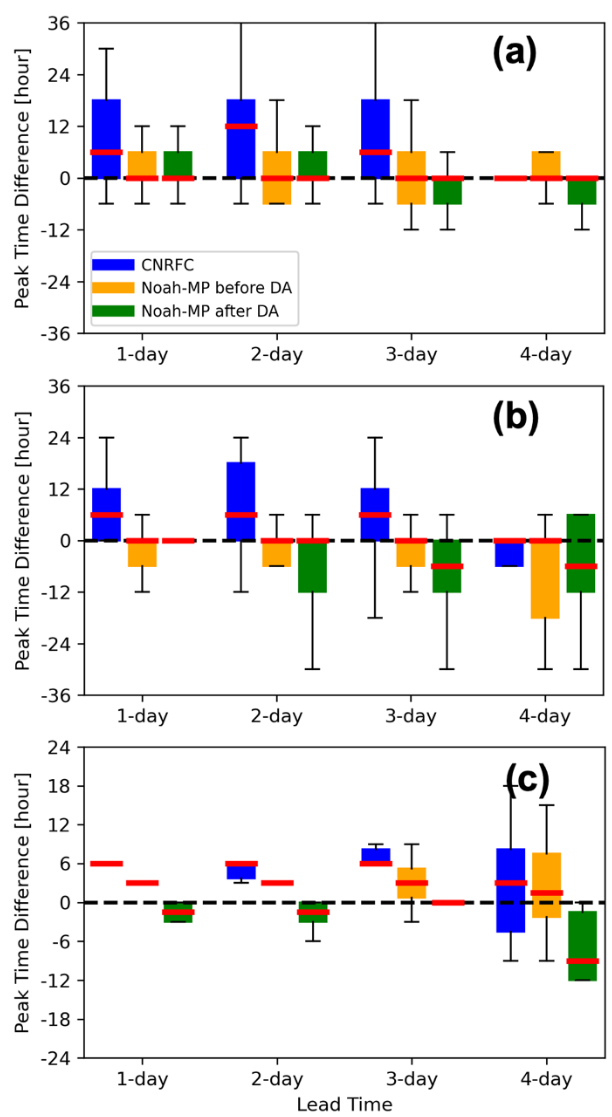


Figure 6. Boxplots of flood peak time differences versus lead days (1, 2, 3, and 4 days) for CNRFC forecasts (blue box), Noah-MP unadjusted forecasts (yellow box), and Noah-MPWM forecasts after QPF bias and initial condition adjustments (green box) in (a) Russian, (b) Yuba-Feather, and (c) Santa Ana River Basins.

Figure 6 shows differences in flood peak timing for 1-, 2-, 3-, and 4-day lead forecasts across the Russian (GUEC1), Yuba-Feather (DCSC1), and Santa Ana (YTLC1) basins. CNRFC forecasts (blue) generally have a late bias (medians mostly less than 12 hours) in forecast peak times. Noah-MP forecasts (yellow; without adjustment) perform better, with median bias closer to zero, though there exist progressively negative timing differences for the Yuba-Feather River Basin at 4-day leads. The most substantial improvement among the forecasts is for Noah-MP forecasts corrected for precipitation bias and with initial condition adjustment (green). These adjustments lead to median peak timing differences near zero and a narrowing of the interquartile range. This trend is consistent across all river basins

(c) Flood hydrograph shape differences

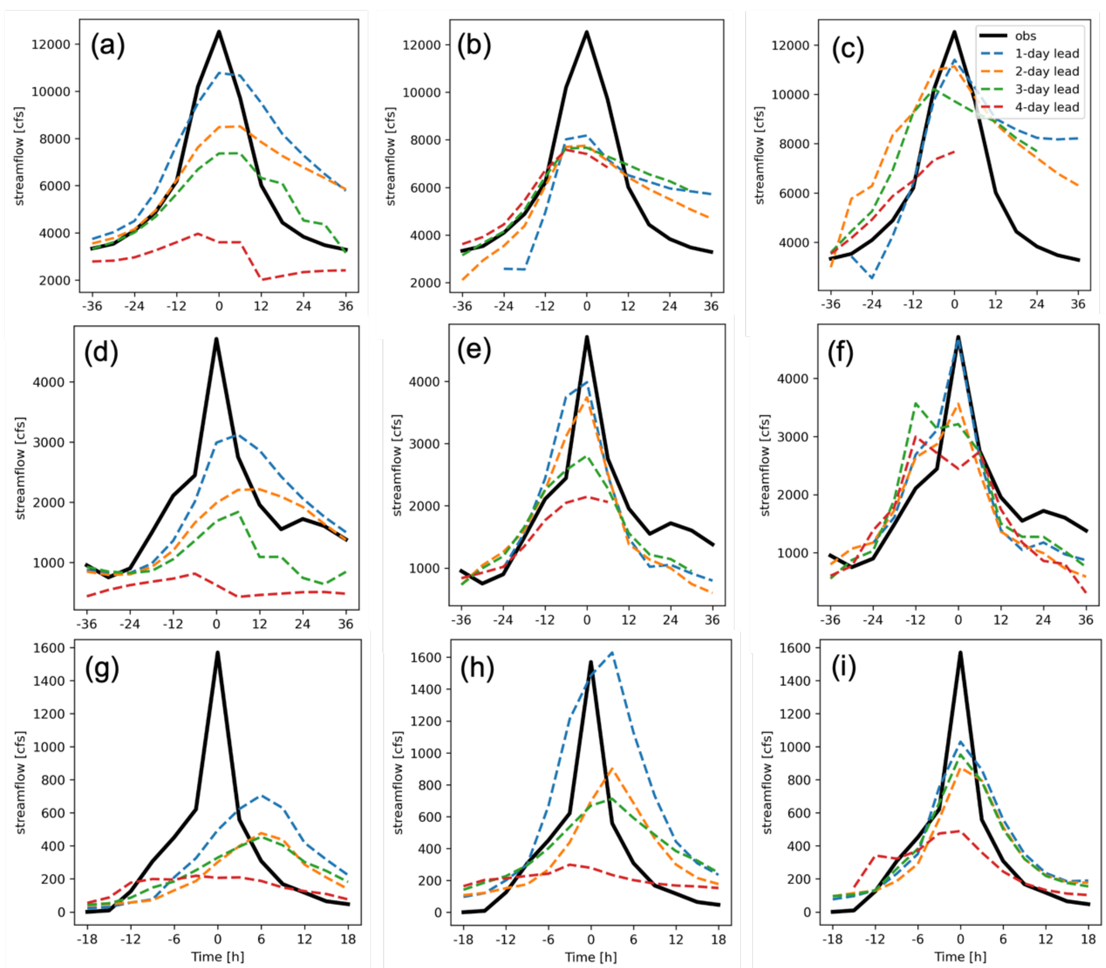


Figure 7. Normalized Flood Hydrographs (Weighted by Flood Peaks) for Russian (a-c), Yuba-Feather (d-f), and Santa-Ana (g-i) Rivers. The flood hydrographs are historical forecasts from CNRFC (first column), Noah-MP unadjusted (second column), and Noah-MP adjustment for QPF bias and soil moisture initial conditions (third column) at different lead days.

Figure 7 shows normalized flood hydrographs for the Russian (GUEC1), Yuba-Feather (DCSC1), and Santa Ana (YTLC1) River Basins, obtained by computing weighted averages (with flood peak discharge as weights) for CNRFC archived forecasts, Noah-MP unadjusted forecasts, and Noah-MP forecasts after QPF bias correction and initial condition adjustment. The shape of the entire hydrograph is important for effective flood forecasting, as it impacts flood preparedness and response strategies. Beyond peak flow predictions, accurately capturing the rise, peak, and recession phases of the hydrograph helps understanding a flood's propagation downstream to ungauged locations. The CNRFC forecasts (Figures 7 (a), (d), and (g)) have a tendency to underestimate peak flows, and to have too long a post-peak recession. The effects of downward bias in peak forecasts (also noted above) tends to result in progressively larger errors in post-peak recessions.

In comparison, the uncorrected Noah-MP forecasts (Figure 8 (b), (e), and (h)) exhibit slightly better overall hydrograph shapes than the CNRFC results, with improved representation of both rising and recession limbs. However, discrepancies remain, particularly as the lead time increases, when post-peak recession errors become progressively larger. The Noah-MP forecasts after data QPF bias correction and initial condition updating (Figure 8 (c), (f), and (i)) show significant improvements. These adjusted forecasts not only capture the flood peaks more accurately but also have forecast hydrograph shapes that are closer to observations.

Discussion

Our results highlight that drier California watersheds (generally to the south in the state) pose greater challenges for calibration due to the infrequency of flood events and the relative scarcity and lower reliability of streamflow records. These factors can limit the effectiveness of model calibration in reducing flood forecast errors. For instance, Russian River station GUEC1 experienced 66 POT3 flood events during the study period (2004–2023), providing a relatively robust dataset for understanding flood characteristics and calibrating the model. However, in the Santa Ana River Basin (station YTLC1), the driest basin, the POT3 definition doesn't really work because there are so few flood events. We eliminated some selected events that were not really peaks, resulting in 25 flood events for analysis. This substantially reduces the size of the calibration data set relative to the other two river basins. The limited number of flood events restricts the ability to capture a comprehensive range of flood characteristics, potentially affecting the accuracy of model parameter calibration and QPF shape adjustment. On the other hand, the same limitations presumably affect performance of the NWS archived forecasts produced using CHPS as well.

In addition to model calibration, our results indicate that systematically unbiasing the QPF and adjusting initial soil moisture before flood events can significantly improve flood forecasting accuracy. These methods could potentially be adopted by the CNRFC in the future.

Currently, the CNRFC does make ad hoc adjustments to both QPF and soil moisture conditions, but these rely largely on individual hydrologists' experience which are difficult to characterize.

Model calibration under changing climate conditions also presents significant challenges. A critical question remains as to whether the parameters calibrated based on historical data will continue to be valid under future climate scenarios. Changes in precipitation patterns, temperature, and other climatic variables can alter watershed hydrology, potentially rendering current model parameters suboptimal or even obsolete. This uncertainty poses limitations for long-term flood forecasting, as models calibrated under present conditions may not accurately reflect the dynamics of future flood events driven by shifting climatic conditions. Addressing this issue may require adaptive calibration strategies that account for projected climate scenarios or periodic re-evaluation of model parameters as new data becomes available. Such approaches could help ensure that models remain robust and reliable in the face of an evolving climate.

Conclusions

In a previous study, Su et al. (2024) found that Noah-MP generally underperformed compared to CNRFC archived forecasts in several basins in coastal California and the Pacific Northwest. In this work, we build on those findings by examining a transect that includes one of Su et al.'s basins (i.e. Russian River Basin) and extends farther south into a drier region, to determine whether Noah-MP can match or exceed CNRFC accuracy under certain conditions. Our investigation reveals that systematic QPF bias correction and improved initial soil moisture estimates (through data assimilation) can significantly enhance Noah-MP's predictive skill. Specifically:

1. In the two northern basins, Noah-MP produced flood reforecasts comparable in accuracy to the archived CNRFC forecasts, both for flood peak magnitudes and timing. Both Noah-MP and CNRFC performed well for lead times of up to about two days but forecast accuracy for CNRFC rapidly declined thereafter. When the lead time exceeded one day, both models tended to underestimate flood peaks, and this underestimation grew with longer lead times. At the longest lead times, Noah-MP typically predicted earlier peaks than observed, while CNRFC usually predicted later peaks than observed.

2. Updating pre-flood soil moisture via data assimilation and correcting precipitation forecast biases resulted in Noah-MP reforecasts becoming more accurate than original Noah-MP reforecasts (in terms of both flood peak magnitudes and timing) for all three river basins, especially at lead times beyond two days.

3. Noah-MP reforecasts generally were less accurate in the Santa Ana (south-most) river basin compared to the northern (generally wetter) Yuba-Feather and Russian River Basins. However, initial soil moisture adjustments and removal of precipitation forecast bias substantially improved forecasts in the semi-arid Santa Ana basin.

Overall, our research suggests the potential of Noah-MP for flood forecasting, particularly with the assistance of suitable model calibration, QPF unbiasing, and data assimilation of discharge observations to update pre-flood soil moisture. For semi-arid basins, model calibration provides some benefits, although these are limited. However, data assimilation is a powerful tool for enhancing flood forecasts in both the two northern basins and the southern semi-arid basin in terms of flood magnitude, peak timing, and hydrograph shape.

Data Availability Statement

All Quantitative Precipitation Estimates (QPE) and Quantitative Precipitation Forecasts (QPF) used in this study are openly available from the California-Nevada River Forecast Center (CNRFC) at https://www.cnrfc.noaa.gov, which also provided the forecast point data.

Acknowledgements

We thank Peter Fickenscher from NWS/CNRFC for providing the CNRFC archived forecasts as well as QPE and QPF used in this paper.

References

Adams, I., Thomas E., and R. L. Dymond, 2019: Possible hydrologic forecasting improvements resulting from advancements in precipitation estimation and forecasting for a real-time flood forecast system in the Ohio River Valley, USA. *Journal of Hydrology*, **579**, 124138, https://doi.org/10.1016/j.jhydrol.2019.124138.

Adams, T. E., 2016: Chapter 10 - Flood Forecasting in the United States NOAA/National Weather Service. *Flood Forecasting*, T.E. Adams and T.C. Pagano, Eds., Academic Press, 249–310.

Bass, B., S. Rahimi, N. Goldenson, A. Hall, J. Norris, and Z. J. Lebow, 2023: Achieving Realistic Runoff in the Western United States with a Land Surface Model Forced by Dynamically Downscaled Meteorology. https://doi.org/10.1175/JHM-D-22-0047.1.

Bournas, A., and E. Baltas, 2021: Increasing the Efficiency of the Sacramento Model on Event Basis in a Mountainous River Basin. *Environ. Process.*, **8**, 943–958, https://doi.org/10.1007/s40710-021-00504-4.

Cai, X., Z.-L. Yang, C. H. David, G.-Y. Niu, and M. Rodell, 2014: Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. *Journal of Geophysical Research: Atmospheres*, **119**, 23–38, https://doi.org/10.1002/2013JD020792.

Chiang, Y.-M., K.-L. Hsu, F.-J. Chang, Y. Hong, and S. Sorooshian, 2007: Merging multiple precipitation sources for flash flood forecasting. *Journal of Hydrology*, **340**, 183–196, https://doi.org/10.1016/j.jhydrol.2007.04.007.

Chifflard, P., J. Kranl, G. zur Strassen, and H. Zepp, 2017: The significance of soil moisture in forecasting characteristics of flood events. A statistical analysis in two nested catchments. *Journal of Hydrology and Hydromechanics*, **66**, 1–11, https://doi.org/10.1515/johh-2017-0037.

CLARKE, A. O., 1996: Estimating Probable Maximum Floods in the Upper Santa Ana Basin, Southern California, From Stream Boulder Size. *Environmental & Engineering Geoscience*, **II**, 165–182, https://doi.org/10.2113/gseegeosci.II.2.165.

Crisan, D., 2001: Particle Filters — A Theoretical Perspective. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. De Freitas, and N. Gordon, Eds., Springer, 17–41.

Daliakopoulos, I. N., and I. K. Tsanis, 2016: Comparison of an artificial neural network and a conceptual rainfall–runoff model in the simulation of ephemeral streamflow. *Hydrological Sciences Journal*, **61**, 2763–2774, https://doi.org/10.1080/02626667.2016.1154151.

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

Djuric, P. M., J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, 2003: Particle filtering. *IEEE Signal Processing Magazine*, **20**, 19–38, https://doi.org/10.1109/MSP.2003.1236770.

Doucet, A., N. de Freitas, and N. Gordon, 2001: An Introduction to Sequential Monte Carlo Methods. *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. De Freitas, and N. Gordon, Eds., Springer, 3–14.

Hammond, M. J., A. S. Chen, S. Djordjević, D. Butler, and O. Mark, 2015: Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, **12**, 14–29, https://doi.org/10.1080/1573062X.2013.857421.

Holtzman, N. M., T. M. Pavelsky, J. S. Cohen, M. L. Wrzesien, and J. D. Herman, 2020: Tailoring WRF and Noah-MP to Improve Process Representation of Sierra Nevada Runoff: Diagnostic Evaluation and Applications. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001832, https://doi.org/10.1029/2019MS001832.

Jain, S. K., and Coauthors, 2018: A Brief review of flood forecasting techniques and their applications. *International Journal of River Basin Management*, **16**, 329–344, https://doi.org/10.1080/15715124.2017.1411920.

Jonkman, S. N., A. Curran, and L. M. Bouwer, 2024: Floods have become less deadly: an analysis of global flood fatalities 1975–2022. *Nat Hazards*, **120**, 6327–6342, https://doi.org/10.1007/s11069-024-06444-0.

Knoben, W. J. M., J. E. Freer, and R. A. Woods, 2019: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, **23**, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019.

Lin, P., M. A. Rajib, Z.-L. Yang, M. Somos-Valenzuela, V. Merwade, D. R. Maidment, Y. Wang, and L. Chen, 2018: Spatiotemporal Evaluation of Simulated Evapotranspiration and Streamflow over Texas Using the WRF-Hydro-RAPID Modeling Framework. *JAWRA Journal of the American Water Resources Association*, **54**, 40–54, https://doi.org/10.1111/1752-1688.12585.

Mascaro, G., A. Hussein, A. Dugger, and D. J. Gochis, 2023: Process-based calibration of WRF-Hydro in a mountainous basin in southwestern U.S. *JAWRA Journal of the American Water Resources Association*, **59**, 49–70, https://doi.org/10.1111/1752-1688.13076.

Miller, A., 2020: A Watershed Event For a Watershed Community: The Development of Flood Control for the Santa Ana River Basin. *History in the Making*, **3**.

Moser, C. L., S. Kroczynski, and K. Hlywiak, Comparison of the SAC-SMA and API-CONT hydrologic models at several Susquehanna River headwater basins.

Neri, A., G. Villarini, and F. Napolitano, 2020: Statistically-based projected changes in the frequency of flood events across the U.S. Midwest. *Journal of Hydrology*, **584**, 124314, https://doi.org/10.1016/j.jhydrol.2019.124314.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, **116**, https://doi.org/10.1029/2010JD015139.

Quenum, G. M. L. D., J. Arnault, N. A. B. Klutse, Z. Zhang, H. Kunstmann, and P. G. Oguntunde, 2022: Potential of the Coupled WRF/WRF-Hydro Modeling System for Flood Forecasting in the Ouémé River (West Africa). *Water*, **14**, 1192, https://doi.org/10.3390/w14081192.

Ralph, F. M., and Coauthors, 2022: Yuba-Feather Forecast Informed Reservoir Operations: Preliminary Viability Assessment.

Salamanca, F., Y. Zhang, M. Barlage, F. Chen, A. Mahalov, and S. Miao, 2018: Evaluation of the WRF-Urban Modeling System Coupled to Noah and Noah-MP Land Surface Models Over a Semiarid Urban Environment. *Journal of Geophysical Research: Atmospheres*, **123**, 2387–2408, https://doi.org/10.1002/2018JD028377.

Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. Mitchell, and F. Chen, 1996: Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of*

Geophysical Research: Atmospheres, **101**, 7461–7475, https://doi.org/10.1029/95JD02892.

Sene, K., 2016: Flash Floods. *Hydrometeorology: Forecasting and Applications*, K. Sene, Ed., Springer International Publishing, 273–312.

Slater, L. J., and G. Villarini, 2016: Recent trends in U.S. flood risk. *Geophysical Research Letters*, **43**, 12,428-12,436, https://doi.org/10.1002/2016GL071199.

Smith, M., and Coauthors, 2013: The distributed model intercomparison project – Phase 2: Experiment design and summary results of the western basin experiments. *Journal of Hydrology*, **507**, 300–329, https://doi.org/10.1016/j.jhydrol.2013.08.040.

Smith, M. B., D.-J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong, 2004: The distributed model intercomparison project (DMIP): motivation and experiment design. *Journal of Hydrology*, **298**, 4–26, https://doi.org/10.1016/j.jhydrol.2004.03.040.

Sofokleous, I., A. Bruggeman, C. Camera, and M. Eliades, 2023: Grid-based calibration of the WRF-Hydro with Noah-MP model with improved groundwater and transpiration process equations. *Journal of Hydrology*, **617**, 128991, https://doi.org/10.1016/j.jhydrol.2022.128991.

Sorooshian, S., Q. Duan, and V. K. Gupta, 1993: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, **29**, 1185–1194, https://doi.org/10.1029/92WR02617.

Sumargo, E., F. Cannon, F. M. Ralph, and B. Henn, 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.

Su, L., R. K. Hartman, D. P. Lettenmaier, and M. Pan, 2024: Improving National Water Model Flood Forecast Skill over Coastal Western U.S. River Basins. Manuscript submitted to J. Hydrometeorol., manuscript number JHM-D-24-0072.

Sun, M., and Coauthors, 2020: Evaluation of Flood Prediction Capability of the WRF-Hydro Model Based on Multiple Forcing Scenarios. *Water*, **12**, 874, https://doi.org/10.3390/w12030874.

Tolson, B. A., and C. A. Shoemaker, 2007: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, **43**, https://doi.org/10.1029/2005WR004723.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1.

Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, **117**, https://doi.org/10.1029/2011JD016048.

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart, 2008: Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resources Research*, **44**, https://doi.org/10.1029/2007WR005940.

Yin, D., Z. G. Xue, D. Bao, A. RafieeiNasab, Y. Huang, M. Morales, and J. C. Warner, 2022: Understanding the role of initial soil moisture and precipitation magnitude in flood forecast using a hydrometeorological modelling system. *Hydrological Processes*, **36**, e14710, https://doi.org/10.1002/hyp.14710.

Zhang, Z., V. Koren, S. Reed, M. Smith, Y. Zhang, F. Moreda, and B. Cosgrove, 2012: SAC-SMA *a priori* parameter differences and their impact on distributed hydrologic model simulations. *Journal of Hydrology*, **420–421**, 216–227, https://doi.org/10.1016/j.jhydrol.2011.12.004.

Zhong, L., Z. Huang, Y. Ma, Y. Fu, M. Chen, M. Ma, and J. Zheng, 2021: Assessments of Weather Research and Forecasting Land Surface Models in Precipitation Simulation Over the Tibetan Plateau. *Earth and Space Science*, **8**, e2020EA001565, https://doi.org/10.1029/2020EA001565.

Zhuo, L., Q. Dai, D. Han, N. Chen, and B. Zhao, 2019: Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application. *Hydrology and Earth System Sciences*, **23**, 4199–4218, https://doi.org/10.5194/hess-23-4199-2019.

Total economic damage caused by natural disasters in the U.S. 1900-2016. *Statista*,. https://www.statista.com/statistics/236530/economic-damage-caused-by-natural-disasters-in-the-us/ (Accessed October 20, 2024).

Chapter 3. Monitoring Reservoir Storage Using SWOT Satellite Observations and a Reservoir Operation Model

This chapter has appeared in print in *Water Resources Research* in March 2026 as doi: 10.1029/2025WR041223.

Abstract

Reservoirs play a critical role in water management, yet comprehensive and real-time observations of reservoir storage remain limited, especially outside the U.S. Observations of two reservoir-related attributes, Water Surface Elevation (WSE) and Surface Area (SA), which can be used to calculate reservoir storage change, are often desynchronized, hindering precise estimation. The recently launched (December 2022) Surface Water and Ocean Topography (SWOT) satellite mission (science observations began August 2023) uses a cutting-edge interferometer to provide global, simultaneous WSE and SA maps of Earth's water bodies, which can be leveraged to estimate reservoir storage change. We evaluate the accuracy of SWOT-based estimates of reservoir storage in comparison with in-situ-based observations for 12 reservoirs in the Western U.S., of which four are in California, four are in the Upper Colorado River Basin (UCRB), and four are in the Columbia River Basin (CRB). Our results show that SWOT produces WSE measurements with less than 20 cm median absolute error (MAE) taken across all 12 reservoirs and 19 months of observations and storage estimates with MAE less than 10%. Model-based reservoir storage estimates (constrained by SWOT observations) can fill temporal gaps accurately and efficiently, even if fewer than one-quarter of SWOT observations are valid. Our results motivate further study of the potential for estimation of reservoir storage

where few or no in-situ observations are available via assimilation of SWOT observations into a reservoir simulation model.

Key Points

- SWOT based WSE observations are within a median absolute error of 20 cm for our larger reservoirs but are considerably larger for the smallest reservoirs.
- SWOT based storage estimates are dominated by WSE accuracy; when well aligned with in-situ WSE, estimation errors typically remain below 10%.
- After assimilation of SWOT observations, model-based storage errors range from 5% to 20%, depending on the number of valid observations.

1. Introduction

Reservoirs play a crucial role in managing water resources by helping to mitigate inter-seasonal and interannual variations in streamflow (Adam et al., 2007; Qian et al., 2016; Vicente-Serrano et al., 2017). Streamflow regulation supports water management functions globally including water supply, flood control, hydroelectric power generation, recreation, and various other water-related services (Biemans et al., 2011; Guan et al., 2020; Hanasaki et al., 2022; Jahandideh-Tehrani et al., 2015; James, 1970; Lu et al., 2020; Ren et al., 2017). Reservoir storage monitoring is therefore important for timely reservoir operations. Better information about reservoir storage can improve system performance under evolving conditions (e.g., storms and droughts) (Benson, 2018). Nonetheless, in situ observations of reservoir storage are often limited or absent (Gao et al., 2012) especially outside the developed world. Although two related attributes of reservoirs, water surface elevation (WSE) and surface area (SA) can be used to calculate reservoir storage change, observations of these variables are often desynchronized in time and have measurement complications, hindering their use for reservoir storage monitoring (Gao, 2015).

Since the 1980s, the development of satellite radar altimetry technology has played an increasingly important role in monitoring water storage of major lakes and reservoirs (Berry et al., 2005; C. M. Birkett, 1994; Charon M. Birkett et al., 2022; Crétaux & Birkett, 2006; Gao, 2015). Originally designed for measuring ocean surface topography, satellite radar altimetry has more recently been applied to estimation of reservoir storage change for very large reservoirs (Gao et al., 2012). However, early applications have been constrained by the fact that radar altimetry as well as LiDAR (Light Detection and Ranging), such as ICESat (Schutz et al., 2005) and ICESat-2 (Abdalati et al., 2010), provide track rather than swath observations with large inter-swath distances, and a requirement for long along-swath track length to produce observations with usable WSE errors (Frappart et al., 2024; Gao et al., 2012; Sulistioadi et al., 2015). In practice, radar altimetry path length requirements (around 10 km) and large across-track spacing (usually hundreds of kilometers) (Birol et al., 2025; Gao et al., 2012) have severely limited the applicability of pre-SWOT altimeters to reservoir storage applications. Nonetheless, until the launch of the Surface Water and Ocean Topography (SWOT) satellite mission in late 2022, track-based satellite altimetry was the most effective method for monitoring WSE changes in large reservoirs from space (Alsdorf et al., 2007; Tang et al., 2009).

SWOT uses a Ka-band (~35 GHz) radar interferometer (KaRIN) to provide global, simultaneous WSE and SA surveys in swaths, rather than tracks, for Earth’s oceans and inland waters. SWOT provides a total swath width of 120 km with unprecedented spatial resolution, offering 100 m raster products and ~22 m pixel cloud (PIXC) products (Kica et al., 2025; JPL D-56411, 2025; JPL D-56416, 2025).

SWOT measurements provide new opportunities to obtain reservoir storage change information for reservoirs with surface areas as small as about 1 km^2 (Hamoudzadeh et al., 2024). The accuracy of SWOT observations has been tested and evaluated using both a SWOT simulator and early SWOT observations. For example, Domeneghetti et al. (2018), using a SWOT simulator, estimated mean elevation errors of 0.1 cm and 21 cm for high and low flows, respectively for the Po river, Italy. Based on real SWOT data, Yu et al. (2024) found that SWOT's global inland water observations are highly consistent with in-situ measurements, showing over 99% agreement across eight regions. Also based on a SWOT simulator, Solander et al. (2016) estimated that SWOT could achieve surface area errors of less than 5% and height errors of less than 15 cm for reservoirs larger than 1 km unless the reservoir exhibited a strong elliptical shape with high aspect ratio oriented parallel to orbit.

As noted above, most SWOT-based studies to date have been based on either SWOT simulators or early SWOT observations, focusing on specific regions or on a global scale. Although several studies have recently validated SWOT's high accuracy for inland waters (Das & Hossain, 2025; Patidar et al., 2025), comprehensive assessments of reservoir storage dynamics, even at the regional level (such as for the Western U.S. (WUS), our topic here) remain limited.

Several factors limit the use of SWOT observations for water management. For example, the revisit time of SWOT is typically from two to 10 days, depending on the locations, which leads to gaps in continuous water body monitoring. Also, the temporal gaps between SWOT observations become larger when satellite data quality preclude use of some observations (e.g. layover issues and heavy rainfall). In contrast, reservoir storage (model) simulations are usually temporally continuous but are often constrained by the lack of reliable storage or elevation observations (especially in the developing world). Under such circumstances, the strategy we pursue here, combining satellite observations with model simulations has considerable appeal.

Here, we explore three motivating questions: (1) How accurate are SWOT-based reservoir WSE and SA observations compared to in-situ observations? (2) How accurate are SWOT-derived reservoir storage estimates compared with those derived from in situ observations? (3) To what extent can model-based reservoir storage estimates fill temporal gaps in SWOT-derived reservoir storage? Insights from this work are intended to improve our understanding of the utility of SWOT-derived reservoir information for global hydrological and water management applications.

Data and Methodology

2.1 Study Area

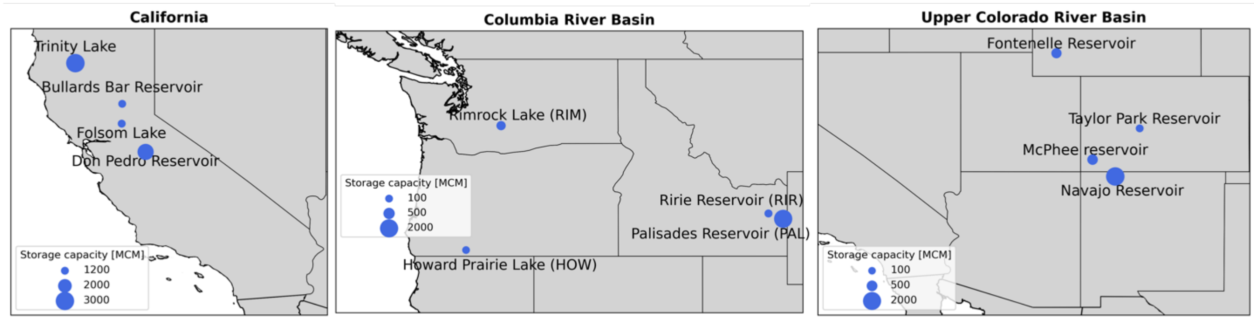


Figure 1. Locations of the 12 selected reservoirs in the Western U.S. Circle sizes represents the storage capacity of each reservoir (see inset).

We selected twelve reservoirs in the WUS (Figure 1). Four (New Bullards Bar Reservoir, Don Pedro Reservoir, Folsom Lake, and Trinity Lake) are in California. Four are in the UCRB (Navajo Reservoir, Taylor Park Reservoir, Fontenelle Reservoir, McPhee Reservoir). Another four are in the CRB (Rimrock Lake, Palisades Reservoir, Howard Prairie Lake, and Ririe Reservoir). Our criteria for selecting the reservoirs were: (1) relatively complete (from ~ 2000 to current) in-situ WSE, storage, and inflow records; (2) no major reservoirs upstream, so the inflow to the reservoirs is relatively natural; (3) relatively complete SWOT WSE and SA observations available; (4) the main purpose of the reservoir is water supply (consistent with the Hanasaki et al. reservoir model; see Section 2.5.2); and (5) there is a relatively large reservoir storage fluctuation (nearly empty to nearly full) from August 2023 to current, so that SWOT's capabilities in capturing reservoir storage changes could be well evaluated. The maximum (active) storage capacities for the twelve reservoirs range from 131 million m^3 (mcm) to 3020 mcm. The specific reservoir information, including abbreviations, storage capacities, regions, main reservoir uses, and sources of in-situ observations are listed in Table 1.

Table 1. Names, sizes, regions, main usages, and sources of in-situ observations of the 12 selected reservoirs in the WUS (California includes the Sacramento and San Joaquin River basin, Columbia is the Columbia River basin above the Dalles, OR, and UCRB is the Upper Colorado River basin (above Lees Ferry, AZ).

Reservoir Name (abbreviation)	Max Capacity (MCM)	Max Surface Area (km^2)	Basin/Region	Main Uses	Sources of in-situ observation
New Bullards Bar Reservoir (BUL)	1,192	18.2	California	water supply, flood control, recreation	Yuba Water Agency
Don Pedro Reservoir (DNP)	2,504	50.0	California	water supply, hydropower	USGS/USACE
Folsom Lake (FOL)	1,205	45.0	California	water supply, recreation	USBR
Trinity Lake (CLE)	3,019	50.9	California	water supply, hydropower	USBR
Navajo Reservoir and Dam (NAV)	2,108	40.9	UCRB	water supply, flood control, recreation	USBR
Taylor Park Reservoir (TAY)	131	8.1	UCRB	water supply, recreation	USBR
Fontenelle Reservoir (FON)	426	30.2	UCRB	water supply, hydropower	USBR
McPhee Reservoir (MCP)	470	15.1	UCRB	water supply, flood control, hydropower	USBR
Rimrock Lake (RIM)	251	8.3	Columbia	water supply, recreation	USBR
Palisades Reservoir (PAL)	1,729	60.2	Columbia	water supply, flood control, hydropower, recreation	USBR

Howard Prairie Lake (HOW)	95	6.0	Columbia	water supply, recreation	USBR
Ririe Reservoir (RIR)	123	6.1	Columbia	water supply, flood control, and recreation	USBR

2.2 SWOT WSE and SA observations

We utilized the "SWOT Level 2 Lake Single-Pass Vector Data Product, Version C (2.0)", which was released in February 2024 (SWOT, 2024) and which we acquired from NASA's Earthdata platform (https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_LakeSP_2.0). We used the SWOT Level-2 High-Resolution Lake/Reservoir Product (L2_HR_LakeSP) to derive water surface elevation (WSE) and surface area (SA) for all reservoirs (we did not use either the PIXC (Pixel Cloud) product or the vector product, which were not available at the time our analysis was conducted). The Version C product (referred to as "L2_HR_LakeSP"), is intended to deliver lake and reservoir-related measurements captured during each continental pass by the high rate (HR) data stream of the SWOT KaRIn instrument. The L2_HR_LakeSP vector product is derived from height, geolocation, and classification data measured by the KaRIn instrument, provided in the "Level 2 KaRIn high-rate water mask pixel cloud product" (L2_HR_PIXC) (SWOT, 2024). The pixel cloud (PIXC) represents the water mask of each water body by classifying geolocated interferogram pixels as water pixels or land pixels (JPL Internal Document, 2018; https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_PIXC_1.1). Another component of the L2_HR_LakeSP product is the shapes and boundaries of each water body, which are extracted by first applying a height-constrained geolocation regularization and then applying the concave hull algorithm to the water edge pixels of the PIXC (JPL Internal Document, 2018; https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_PIXC_1.1). After determining boundaries, we computed the attributes of each water body, including WSE and SA as the average of all water pixels of PIXC within the boundary. In short, while LakeSP is derived from PIXC, our analysis relied solely on the final LakeSP product and did not use PIXC directly.

Initial SWOT data were associated with a one-day repeat orbit for the mission's so-called “calibration” or “fast-sampling” phase, which concluded in early July 2023. SWOT then transitioned to a 21-day repeat orbit in August 2023 which commenced its “science” phase (JPL Internal Document, 2018; https://podaac.jpl.nasa.gov/dataset/-SWOT_L2_HR_PIXC_1.1). Here, we used only data from the science phase, which are fully calibrated for scientific applications. At low to mid latitudes, most locations are observed once or twice per 21-day repeat cycle, often with an approximate 10.5-day interval if both ascending and descending passes occur. At higher latitudes, due to orbit convergence, observation frequency increases and the interval between revisits is typically shorter than 10.5 days.

The measured attributes we used, WSE and SA, are provided in a feature dataset that covers the full swath for each continent and each overpass. The reservoirs we selected are identified in the Prior Lake Database (PLD; Wang et al., 2023) so that each reservoir is assigned with a prior identifier that facilitates data organization. As mentioned above, these data are primarily generated for inland and coastal hydrological surfaces, regulated by the reloadable KaRIn HR mask, and is provided in the Environmental Systems Research Institute (ESRI) Shapefile format.

We filtered out SWOT observations with poor quality, including: (1) fraction of prior lake covered by each observed lake (overlap) > 90%, so that we could obtain relatively complete SA observations; (2) the "quality flag" (quality_f) for which the observation was not flagged as either 0 (good) or 1 (suspect), that is, we excluded all observations with flags 2 (degraded) and 3 (bad) entirely. To ensure a sufficient number of available SWOT observations, we included both quality_flag values 0 and 1, which together account for approximately 50% of the total observations. Although labeled as "suspect," flag 1 observations are not considered erroneous and may still represent physically reasonable water surfaces, but with higher uncertainty according to SWOT documentation (SWOT, 2024). We retained observations with a quality_flag of 1 because those with a flag of 0 were too limited in number (only 10–20% of all observations) to support our analyses. In addition, many flag 1 observation showed temporally consistent patterns with reservoir dynamics, indicating that they were usable for this study.

For WSE, sometimes there exist systematic biases between in-situ and SWOT observations because of datum differences. We corrected the systematic biases in SWOT WSE observations by subtracting the average differences, which we calculated based on all available observations except outliers (difference > 0.6m).

2.3 In-situ WSE, storage, and inflow observations

We focus on reservoir storage rather than storage change. To ensure consistency, we converted storage change values into absolute storage using a (single) known reference storage value for each reservoir (see details in section 2.5.1). All subsequent analyses and comparisons are thus presented in terms of storage. We processed in-situ observations for the validation of SWOT observations. The observed attributes include WSE, reservoir inflows, and observation-derived reservoir storage. The observation-derived storage is usually based on an elevation sensor and prior reservoir bathymetry (contour maps). The twelve reservoirs have relatively complete WSE and storage observations, which we leveraged as "ground-truth" for purposes of evaluating SWOT-derived reservoir storage. The inflow observations can be fed into models for the simulation of reservoir storage. Table 1 summarizes in-situ data sources.

2.4 Derived in-situ-based SA estimates

We processed the daily in-situ WSE and storage observations from 2017 Jan to 2025 Feb and then fitted quadratic equations to the WSE-storage curves (Figure 2). We evaluated SWOT's WSE performance relative to in-situ observations, as shown in Figure 3. The derivatives of these relationships are the estimated WSE-SA curves (Figure 4). We used SA values from the WSE-SA curves as surrogates for in situ observations of SA. We note that the framework we used could be extended to ungauged regions by leveraging remotely sensed WSE-storage and WSE-SA relationships derived from multi-temporal satellite imagery (Avisse et al., 2017; Busker et al., 2019; Gao, 2015). Such approaches have shown promise in constructing or updating reservoir storage curves, which may complement SWOT observations in regions lacking in-situ data.

2.5 Estimation of reservoir storage

2.5.1 SWOT-based reservoir storage estimates

We obtained time series of SWOT-derived reservoir storage by calculating the storage differences relative to a baseline storage (Eqs. 1 and 2, Figure 2(a)). The baseline storage and the baseline WSE are from in-situ observations (Table 1).

$$\Delta S = (WSE_i - WSE_b) * SA_i \quad (1)$$

$$S_i = S_b + \Delta S \quad (2)$$

where ΔS is the storage difference between the SWOT storage estimate and baseline reservoir storage S_b; WSE_i is the SWOT WSE observation at time i; WSE_b is the WSE corresponding to the baseline storage; SA_i is the SWOT SA observation at time i; and S_i is the SWOT-based reservoir storage estimates (our target).

We obtained the baseline storage from in-situ elevation-storage curves (Figure 2 (a)). The elevation-storage curves were plotted using the in-situ WSE and storage observations from December 2017 to December 2024. When there is more than one curve shown (which means there may exist more than one relationship between WSE and storage), we shrank the period into a more recent one to make sure there was only one single curve. The reasons for multiple curves include (1) sediment changes altering WSE-storage relationships; (2) observation inconsistence; (3) observation errors. We first selected a baseline reservoir WSE at around 50% of the full WSE range for each reservoir. We fit a quadratic equation for each in-situ WSE-storage curve (Figure 2 (b)), and the baseline reservoir storage was calculated from the baseline WSE and the fitted quadratic function. In practice, the baseline WSE and the corresponding baseline reservoir storage can come from any reservoir records and documents.

After determining the baseline WSE and reservoir storage, we estimated reservoir storage at each SWOT observation time for which there was a valid SWOT WSE and SA observation, based on Eqs. (1) and (2). Thus, we obtain the time series of SWOT-based reservoir storage estimates during the SWOT observation period (August 2023 – February 2025).

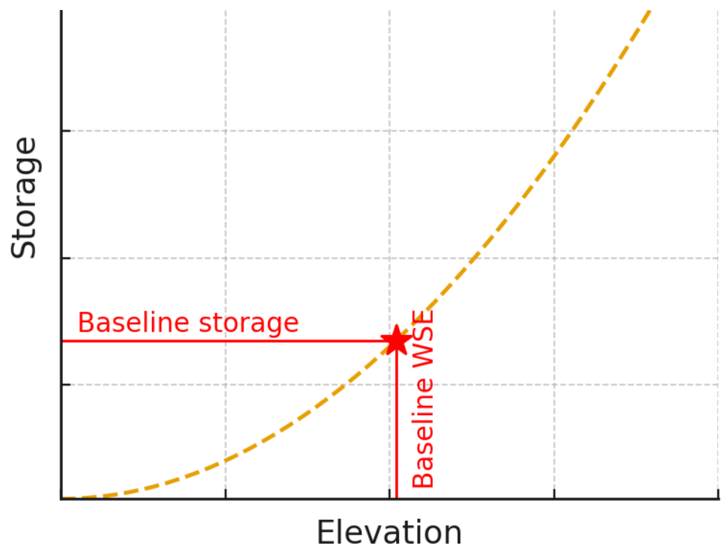


Figure 2(a). schematic illustration of baseline WSE and storage

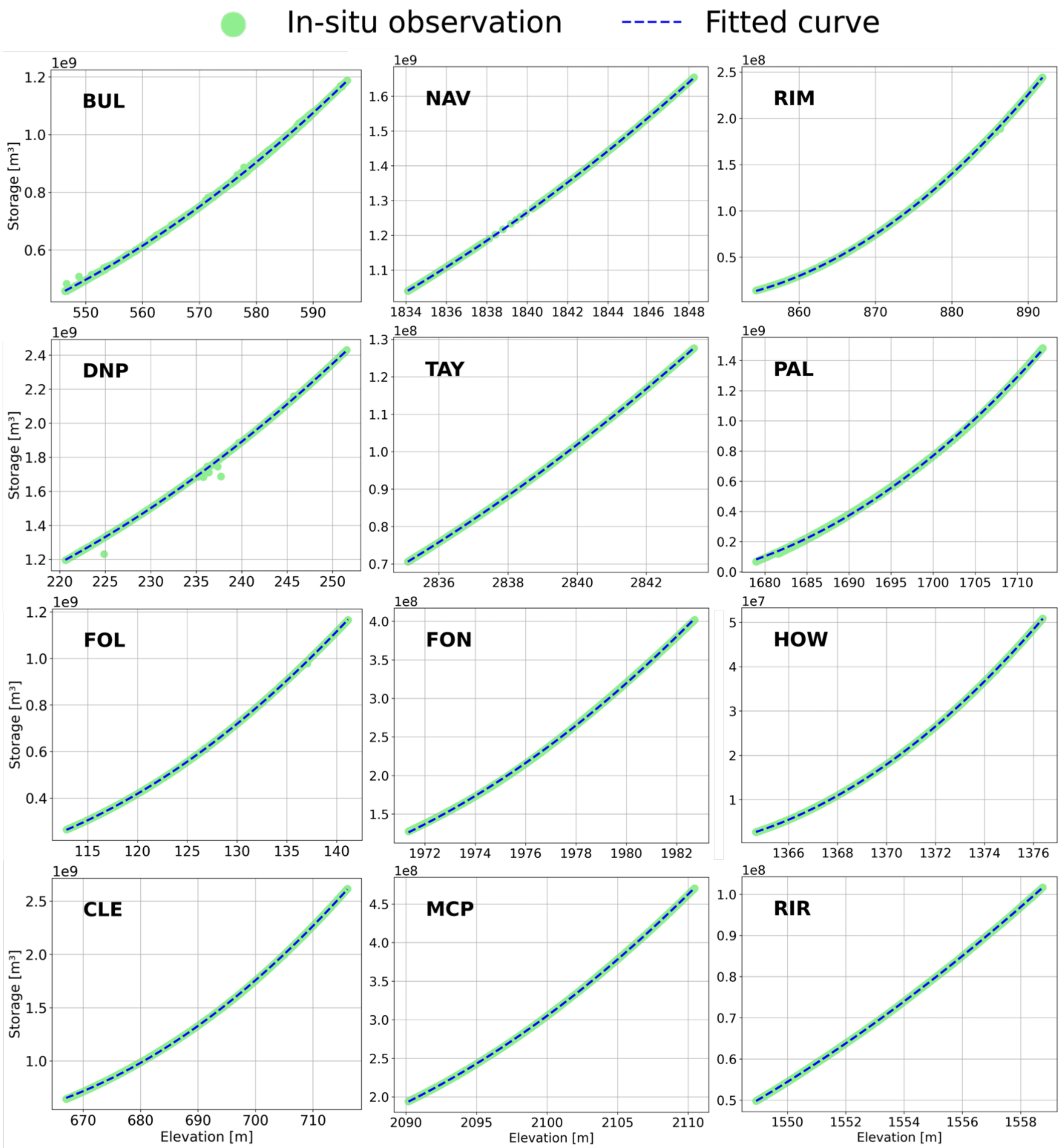


Figure 2(b). in-situ WSE-storage curves, and corresponding fitted functions for the 12 WUS reservoirs in California (first column), the UCRB (second column), and CRB (third column).

2.5.2 Model-based reservoir storage estimates (Hanasaki algorithm)

In addition to satellite and in-situ observations, reservoir operation models can also be used to estimate reservoir storage. On continental and global scales, the most commonly used reservoir operation models simulate reservoir impacts on downstream river discharge to help regulate reservoir outflow and storage (P. Döll et al., 2009a; Haddeland et al., 2006; Hanasaki et al., 2006). For reservoirs whose primary operating purpose is water supply, we found the Hanasaki et al. (2006) model to be most useful in addressing our motivating questions (section 1).

Here, we used the Hanasaki algorithm to perform reservoir storage estimates from 2009 to 2025 and evaluated model performance in comparison with in-situ-derived storage (using the same in-situ sources listed in Table 1). More specifically, the primary inputs to the Hanasaki algorithm are inflow and water demand time series, often on a monthly basis. The model adjusts the demand term according to the reservoir's current storage level. The Hanasaki algorithm scales the demand according to the reservoir's current (at the regulation step) storage level. It is primarily applied to reservoirs with water supply as their main purpose. Its performance—compared to the Döll natural lake model (P. Döll et al., 2009) —has been evaluated on a global scale by Vanderkelen et al., 2022.

We used the Hanasaki reservoir operation algorithm because it provides a simple yet physically interpretable representation of reservoir release behavior and has been widely applied in large-scale hydrologic modeling (e.g., Hosseini-Moghari & Döll, 2025; Sadki et al., 2023; Vanderkelen et al., 2022b). Its modest data requirements also make it suitable for reservoirs with limited information. We took the values of the model parameters from Gharari et al., 2024. We took maximum capacity (MC) data from reservoir documents (see Table 1). Among all the parameters, the partial of dead storage to MC (P_{dead}), partial of active storage to MC (P_a), monthly inflow, and monthly demand are the most important. We determined parameters by manual calibration using the historical inflow data from Jan 2009 to July 2023 (immediately before the SWOT period). After determining the Hanasaki parameters, we ran the Hanasaki model using the observed inflow to each reservoir from 2007 to 2025 (using 2007-2009 as the spin-up period). The outputs are simulated reservoir storage and outflows.

The results of our Hanasaki-based (model) reservoir storage estimates are generally consistent with in-situ measurements (see section 3.4). Temporally continuous model-based reservoir storage simulations provide the possibility for infilling temporal gaps in SWOT observations, as we describe in the next section.

2.5.3 SWOT Gap-Filling Using the Hanasaki Algorithm

We selected one reservoir for each region (DNP in California, NAV in UCRB, and RIM in Columbia) to perform a case study to investigate the performance of the Hanasaki model in filling temporal gaps in SWOT reservoir storage estimates. The three reservoirs were selected because they showed relatively good performance for both SWOT storage estimates and Hanasaki storage simulations relative to in situ-derived reservoir storage. SWOT observations are theoretically available at intervals of 10.5d or less in most cases, but only around 50% of the theoretically available observations were available in practice due to poor observations quality (section 2.2). Based on the available good-quality SWOT observations, we designed a synthetic experiment for the high storage volume period (April to September 2024). We defined four scenarios in which the number of observations retained was none, less than a quarter, half, or all of the available “good” SWOT observations (the standard of available “good” SWOT observations is described in section 2.2). These scenarios simulate different levels of data availability by selectively removing observations between the synthetic “first” and “last” SWOT overpasses. We initialized the Hanasaki algorithm with the beginning-of-gap SWOT storage estimate and then scale (by multiplying a constant smaller or larger than 1) the inflow to force the end-of-gap Hanasaki estimate to match the SWOT end-of-gap storage. This procedure assures that the beginning and end of gap Hanasaki storage estimates match those from SWOT. By exploring whether the storage simulation in-between matches the in-situ storage observations, we tested the ability of the Hanasaki algorithm to produce serially realistic storage estimates based on SWOT, but accounting for gaps in the SWOT record as indicated by SWOT data quality flags.

Results

3.1 SWOT WSE observations

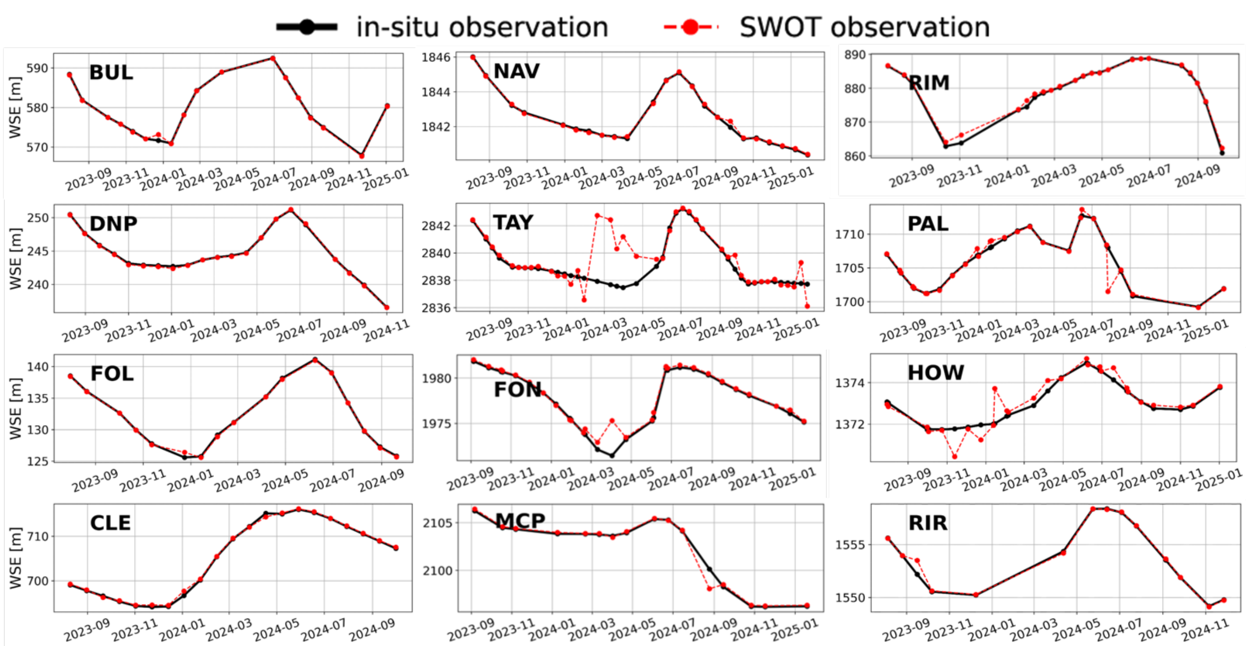


Figure 3. Time series of SWOT WSE against in-situ observations in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column).

Figure 3 shows the time series of SWOT WSE observations compared with in-situ observations in all twelve reservoirs. The reservoirs we selected all had relatively large seasonal storage fluctuations during the SWOT period, hence they provide a good basis for evaluating SWOT WSE (and storage) estimates over realistic ranges of WSE and storage for each of the reservoirs.

For the California reservoirs, the SWOT WSE series generally align well with in-situ observations. Statistically, the median absolute errors (MAE) for BUL, DNP, FOL, and CLE are 6.2, 5.3, 5.9, and 6.0 cm (Table 2), which are small compared to the seasonal variations in WSE. The 90th percentile absolute errors (PAE) overall were < 30 cm for the California reservoirs.

In the UCRB, systematic biases in SWOT WSE observations are larger than in California, possibly due in part to datum differences and uncorrected biases in KaRIN instrument. We removed these biases by adding or subtracting a constant to minimize the MAE (we performed the same adjustments in for the CRB reservoirs). After these adjustments (which are incorporated in the SWOT WSEs shown in Figure 3) we found that two factors seemed to influence the accuracy of SWOT WSE observations in the UCRB: (1) reservoir size and (2) reservoir water level. In general, our results suggest that the larger reservoirs have lower WSE error. For example, NAV, which is the largest reservoir in the UCRB, has the smallest MAE among the UCRB reservoirs (5.7 cm). In contrast, TAY is the smallest UCRB reservoir and has some large errors (90th PAE is 184.6) especially when the reservoir water level is low. Overall, the MAEs are 5.7, 14.6, 15.6, and 10.9 cm in NAV, TAY, FON, and MCP. The larger errors mostly correspond to fewer pixels used for WSE calculation when reservoir surface areas are small. For fewer pixels, the number averaged is smaller, but also the influence of edge pixels is greater. In addition, at lower water levels, reservoirs shores are more exposed to mud etc., which can be confounded with water. For example, in TAY (the smallest reservoir among the four selected in the UCRB), at the lowest WSE values, the fraction of non-water exposed bank area is ~ 22.5%. Furthermore, all reservoirs exhibit substantial exposed bank areas at the lowest WSE: TAY (~22.5%), NAV (~23.6%), MCP (~33.3%), and FON (~39.5%), suggesting that edge effects and misclassification are a general issue at low surface areas.

SWOT observations in the CRB show patterns that are somewhat similar to those in the UCRB, that is, larger errors occur for smaller reservoirs and lower reservoir water levels. The MAEs are 17.9, 10.9, 14.5, and 4.9 cm in RIM, PAL, HOW, and RIR. The 90th percent errors for the four reservoirs in the CRB are shown in Table 2. Smaller reservoirs tend to have larger errors in SWOT-derived WSE. For example, within the twelve reservoirs in the Western U.S., HOW has the smallest storage capacity and has relatively large uncertainties in its SWOT WSE estimates (both in MAE and 90th-PAE errors).

Table 2. MAE and 90th PAE of SWOT WSE observations in the 12 reservoirs.

Region	Reservoir	WSE MAE [cm]	WSE 90th PAE [cm]
California	BUL	6.2	15.6
	DNP	5.3	14.4
	FOL	5.9	16.3
	CLE	6.0	28.7

UCRB	NAV	5.7	11.3
	TAY	14.6	184.6
	FON	15.6	59.2
	MCP	10.9	21.4
CRB	RIM	17.9	122.3
	PAL	10.9	88.4
	HOW	14.5	69.4
	RIR	4.9	15

3.2 SWOT SA observations

We explored the performance of SWOT SA observations as well. Comparison with in-situ SA observations is more complicated than in the case of WSE as direct in situ SA observations generally aren't available. Instead, we took surrogate in-situ SA estimates from in situ WSE-SA relationships. The WSE-SA relationships (Figure 4) are the derivative function of the WSE-storage curve (Figure 2). We also fitted WSE-storage curves to the in-situ observations, whose sources are specified Table 1. Details as to how we obtained the surrogate in-situ SA estimates are described in section 2.4.

Figure 4 shows that for the California reservoirs, there generally is better agreement between SWOT and in-situ SA surrogate observations (Figure 4 first column) than for the other regions. Furthermore, for FOL and CLE, SWOT tends to overestimate SA especially for low WSEs (e.g., lowest 30% in SWOT WSE). For BUL and DNP, there is a slight overestimate for the highest WSE values.

The SWOT SA observations are worse in the UCRB River and CRBs (Figure 4 second and third columns). For NAV and FON, there is a distinct overestimate in SWOT SA observations – SWOT observed 4% to 43% larger SA compared to the in-situ SA surrogates in

NAV with up to 40% overestimate in FON. For TAY, at high WSE (upper 50%), SWOT observations systematically underestimate SA. For MCP, there are a few outliers for low WSE with errors exceeding 60%.

The same SWOT SA error patterns emerge in the CRB (Figure 4 third column). For RIM, the in-situ surrogate WSE-SA curve has a slope ~100000 (m-m^2). However, all SWOT SA observations for RIM are around 9.7 km^2 with no slope. For PAL, the SWOT WSE-SA pairs have a similar slope as for the in-situ surrogate, but SWOT SA observations nonetheless have errors in the range -20% to +20%. For HOW and RIR, SWOT SA values generally have a larger bias than for the other sites with errors exceeding 50%. The comparison between SWOT and in-situ WSE-SA curves shows an overall good match in California but relatively worse alignment in the UCRB and CRBs.

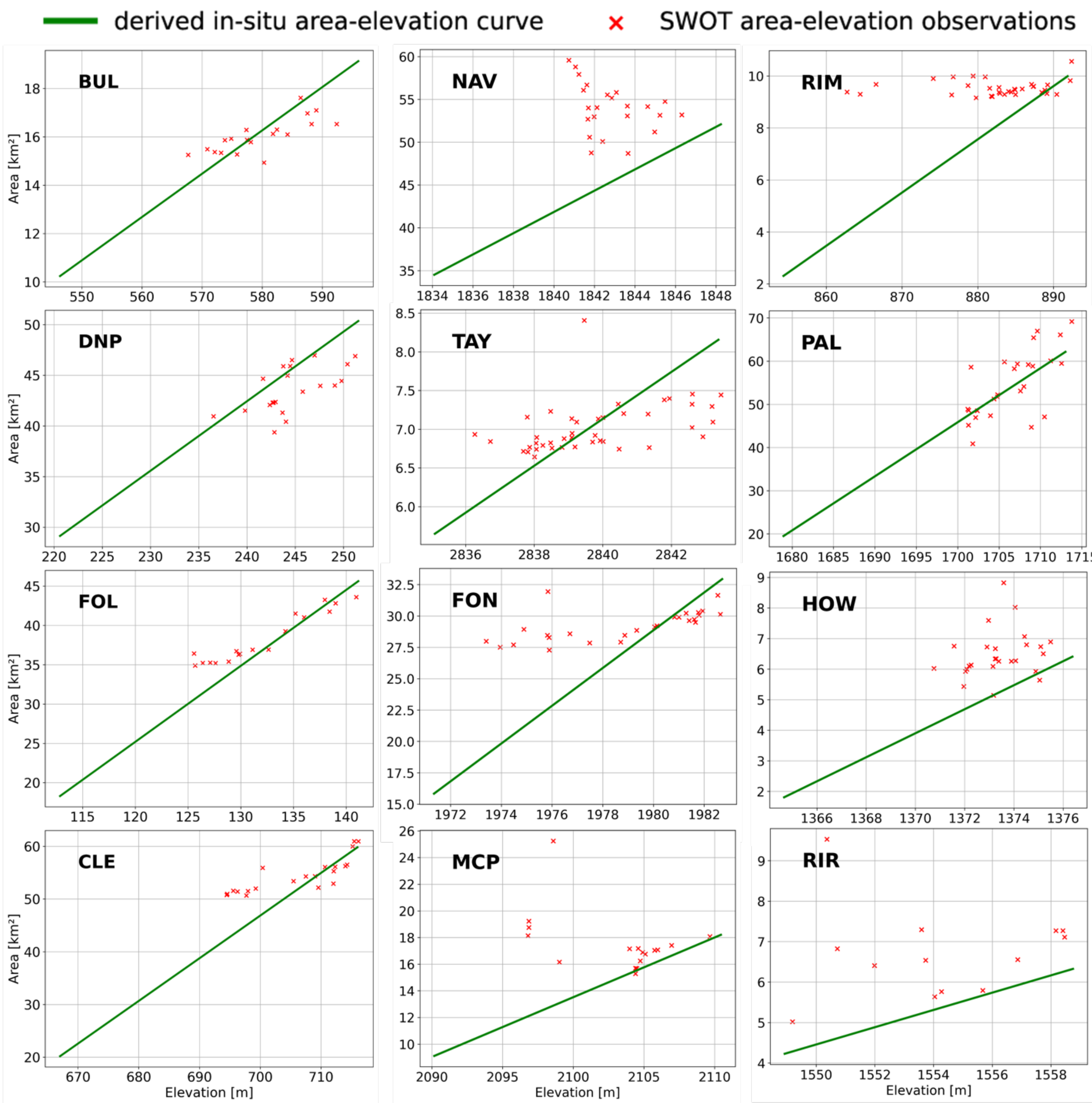


Figure 4. derived in-situ WSE-SA curves compared with SWOT-derived WSE-SA pairs over the WUS in California (first column), the UCRB (second column), and CRB (third column).

We evaluated SWOT SA time series in comparison with surrogate in-situ observations (Figure 5). To obtain the time series of in-situ surrogate SA values, we first extracted in-situ WSE values from Figure 3, and then used the WSE values at the SWOT observation times to derive the corresponding SA values from the in-situ elevation–area curves shown in Figure 4. The time series of SA are plotted in Figure 5.

For the California reservoirs (Figure 5 first column), the overall seasonal variations in in-situ and SWOT SA match somewhat. However, for all the four reservoirs, SWOT has an overall SA underestimation in summer (when the reservoir storage is relatively full) and overestimation in fall, winter, and spring (when the reservoir is relatively empty). Compared to the in-situ surrogate SA series, the SWOT SA series does not show nearly as large seasonal variations, nor are the variations as smooth. A possible reason is that the in-situ surrogate SA series are derived from WSE-SA curves, which assume linear relationships between the observed in-situ WSE and the corresponding surrogate SA.

In the UCRB, the SWOT SA time series have larger variability and outliers than for California (Figure 5 second column), likely because the smaller SA values result in larger errors, and more difficulty in detecting systematic variations in SA (e.g., in late summer and fall). For NAV and FON, SA is persistently overestimated by SWOT. In fact, SWOT SA observations are more or less constant and show no obvious seasonal fluctuations. For TAY, the SWOT observed SA values are too small in summer; in MCP, higher outliers in SWOT SA are apparent from September 2024 to present. The same situation occurs in the CRB (Figure 5 third column), with either more or less constant SWOT SA observations (RIM), or larger observation outliers with little apparent seasonal fluctuations (PAL, HOW, and RIR).

As our in-situ SA surrogates are inferred from the derived elevation-area curves (which also are estimates rather than observations), there clearly exist the potential for biases and/or errors in our benchmark. However, the SWOT SA observations lack smooth changes in their time series which must be present in fact (clearly, SA must have a monotonic relationship with SA), and the outliers are unrealistic. In general, the quality of SWOT surface area (SA) observations tends to degrade at relatively low water surface elevations (WSEs).

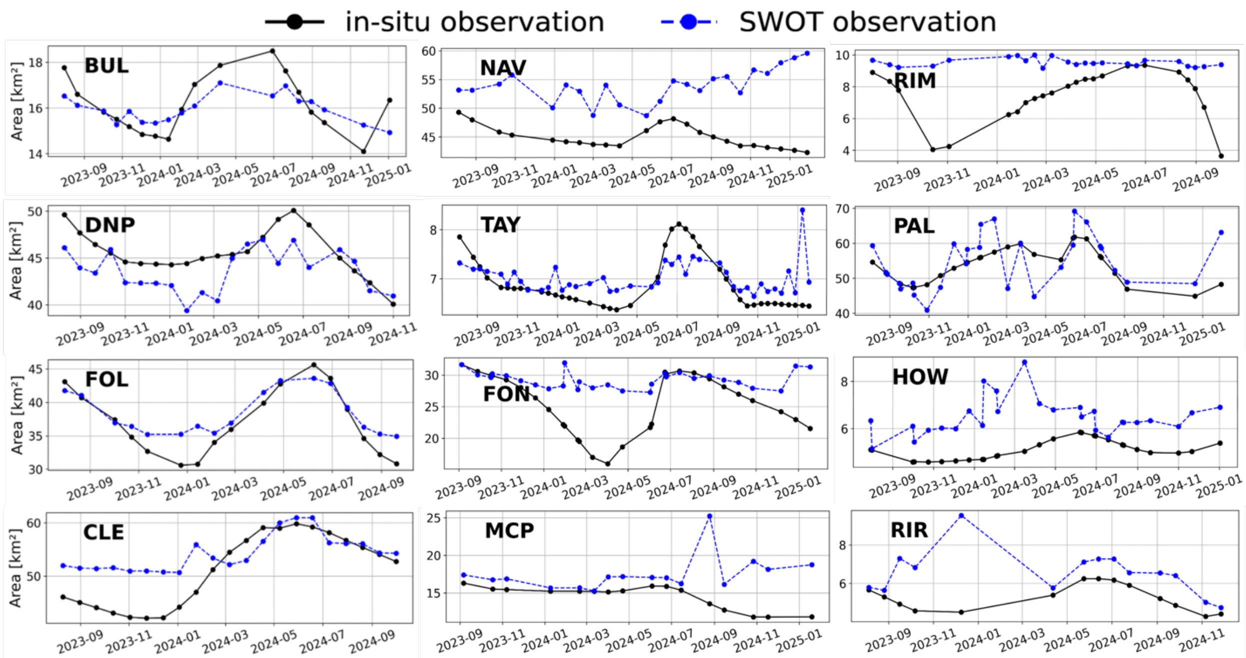


Figure 5. Time series of SWOT SA against derived in-situ SA in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column).

3.3 SWOT storage estimates

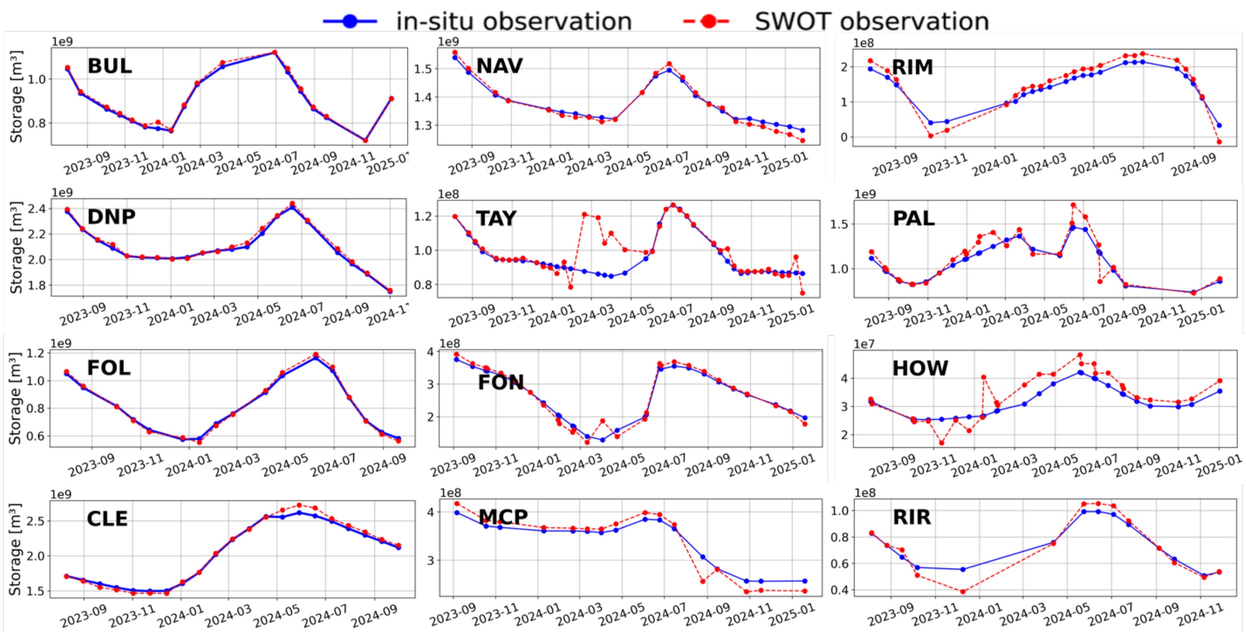


Figure 6. Time series of SWOT reservoir storage against in-situ storage observations in 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column).

Using SWOT-derived WSE and SA, which were adjusted at a single point to match the baseline WSE and storage (see Section 2.5.1 for details on the selection of baseline pairs), we derived reservoir storage time series based on SWOT observations (Figure 6; see Section 2.5.1 for methodological details). In the remainder of this section, we refer to this time series as "SWOT storage". In general, errors of SWOT storage series are controlled mostly by SWOT WSE errors. Even though SWOT WSE and SA observations are discontinuous and not generally smooth, the SWOT storage values nonetheless mostly reflect coherent seasonal fluctuations. Reasons that SWOT storage errors are dominated by SWOT WSE errors include: (1) The fluctuation range for SWOT SA is smaller than that of SWOT WSE. For example, for BUL, the total SA fluctuation range is around 10% compared to its average, while total WSE fluctuation range is around 30%. (2) during the process of reservoir storage estimation, the estimated SWOT storage is more sensitive to WSE than SA (not due to the larger fluctuation range of WSE). To further investigate, we performed a simple sensitivity test based on equation (1) and (2) applied to BUL. For BUL, WSE_b is 573 m, and S_b is about $0.8 * 10^9$ m^3. The average of the absolute relative error for the SWOT reservoir storage at BUL is 0.97%. As a sensitivity test, we added a 1% bias to all of the original WSE_i values which led to a very large relative storage error of 15.96%. Keeping the WSE_i as original, we also added 10% biases to SA_i, which resulted in a final average absolute relative storage error of only 2.1%. Following the results in sections 3.1 and 3.2, SWOT WSE generally performs much better than SWOT SA. Thus, we are able to produce storage estimates that are quite good based on SWOT WSE and SA observations, even given the substantial SWOT SA errors. We note that, for each reservoir, our selected baseline is near the median value over the WSE range. Based on equation (1) and (2), if the SWOT WSEs are smaller than the baseline WSE, then the positive biases in SWOT SA will cause a negative bias in SWOT storage estimates; if the SWOT WSE is larger than the baseline WSE, then the positive biases in SWOT SA will cause a positive bias in SWOT storage estimates. One important caveat is that our example (BUL) is a California reservoir with quite good SWOT WSE values. As noted in section 3.1, some WSE errors for the smaller UCRB and CRB reservoirs are much larger than for the California reservoirs, which will lead to larger storage errors.

Among the California reservoirs, for BUL, DNP, and FOL, there is an overall good match between the SWOT and in-situ storage estimates, with absolute relative errors generally less than 5% (average absolute relative errors 0.97%, 0.66%, and 1.84% for BUL, DNP, and FOL, respectively). For CLE, SWOT-estimated reservoir storage tends to overestimate in situ values (average about 4.5% positive relative errors) due to the higher SWOT SA observations relative to in-situ observations.

In the UCRB, for NAV, FON, and MCP, the largest storage errors (all overestimates) occur in summer when the reservoir storage is high, mainly due to SWOT SA biases. In the smallest reservoir (TAY), reservoir storage errors are quite large (around 30%) due mostly to errors in SWOT WSE (Figure 3).

In the CRB, even though SWOT storage estimates are less sensitive to SWOT SA errors than SWOT WSE errors, the large errors in SWOT SA nonetheless bias the storage estimates. For RIM, though there is a large positive bias in SWOT SA at the end of 2023, the positive biases in SWOT SA and negative biases in SWOT WSE offset each other, so SWOT performs well in SWOT storage estimates at this location. For PAL, HOW and RIR, errors in SWOT SA are the dominant factor in overall reservoir storage estimates.

Overall, SWOT-derived reservoir storage estimates performed better in California compared to the UCRB and CRB. This mostly is due to the generally larger size of California reservoirs. The reservoirs in the UCRB and CRB are relatively smaller in surface area (i.e. TAY, HOW, RIR; see Table 1), which increases the influence of edge pixels in SWOT SA observations. This is an issue not only for small but also for narrow reservoirs (Durand et al., 2023a). In addition, the complex terrain in these regions can introduce radar-specific issues such as layover and shadowing, which distort the interferometric signal (Durand et al., 2020). More specifically, CLE, TAY, and FON are the three reservoirs most impacted by layover and shadowing issues. Environmental conditions also play a role; heavy precipitation, dense vegetation, and spatially variable surface moisture are more common in the Pacific Northwest during winter, potentially degrading radar backscatter quality during this period (Heffernan & Strimbu, 2021). In contrast, California benefits from a denser network of in-situ gauge stations, which provides more opportunities for spatial calibration, potentially enhancing the accuracy and interpretability of SWOT observations. These combined factors likely explain the regional differences in SWOT uncertainty.

3.4 Hanasaki storage estimates

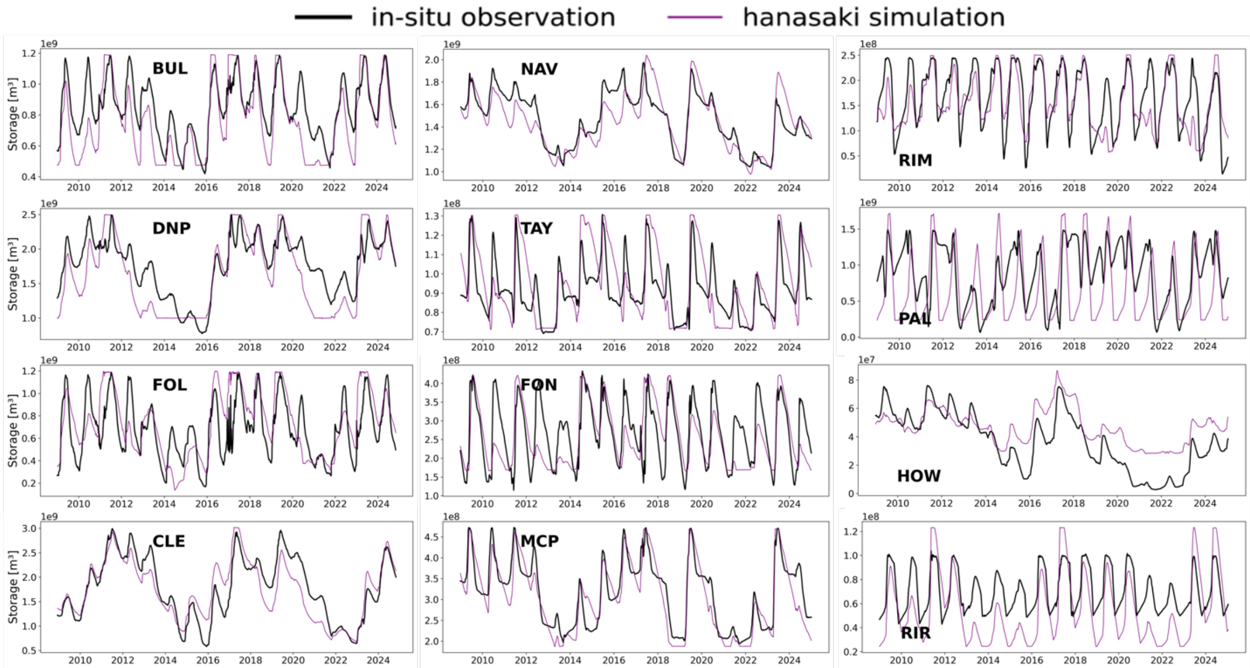


Figure 7. Hanasaki-based storage estimates compared with in-situ storage observations from 2009 – 2025 for the 12 reservoirs over the WUS in California (first column), the UCRB (second column), and CRB (third column).

Model-based reservoir storage simulations show consistent trends and patterns compared with the in-situ storage measurements over the longer (2008 to 2024) period, although the accuracy varies considerably for specific years (Figure 7). The Kling–Gupta efficiency (KGE; Gupta et al., 2009) is a widely used metric for hydrological model evaluation, ranging from –∞ to 1, with 1 indicating perfect agreement between simulated and observed values. We computed the R^2 and KGE values between simulated and measured reservoir storage (shown in Table 3.) Except for a few low values, the overall R^2 values (mostly larger than 0.5) suggest the potential for using model-derived storage to augment SWOT estimates when necessary. The KGE values show that in general, the Hanasaki algorithm shows better performance in California and the

UCRB relative to the CRB. In the simulations shown in Figure 7, we used observed streamflow as input, so most of the error comes from differences in the model's assumptions as to how the reservoirs are operated.

Table 3. Statistical performance (R^2 and KGE) comparing simulated and observed reservoir storage.

Region	Reservoir	R^2	KGE
California	BUL	0.68	0.67
	DNP	0.7	0.66
	FOL	0.48	0.66
	CLE	0.74	0.83
UCRB	NAV	0.72	0.85
	TAY	0.5	0.52
	FON	0.44	0.66
	MCP	0.85	0.89
CRB	RIM	0.52	0.69
	PAL	0.12	0.3
	HOW	0.7	0.5
	RIR	0.61	0.4

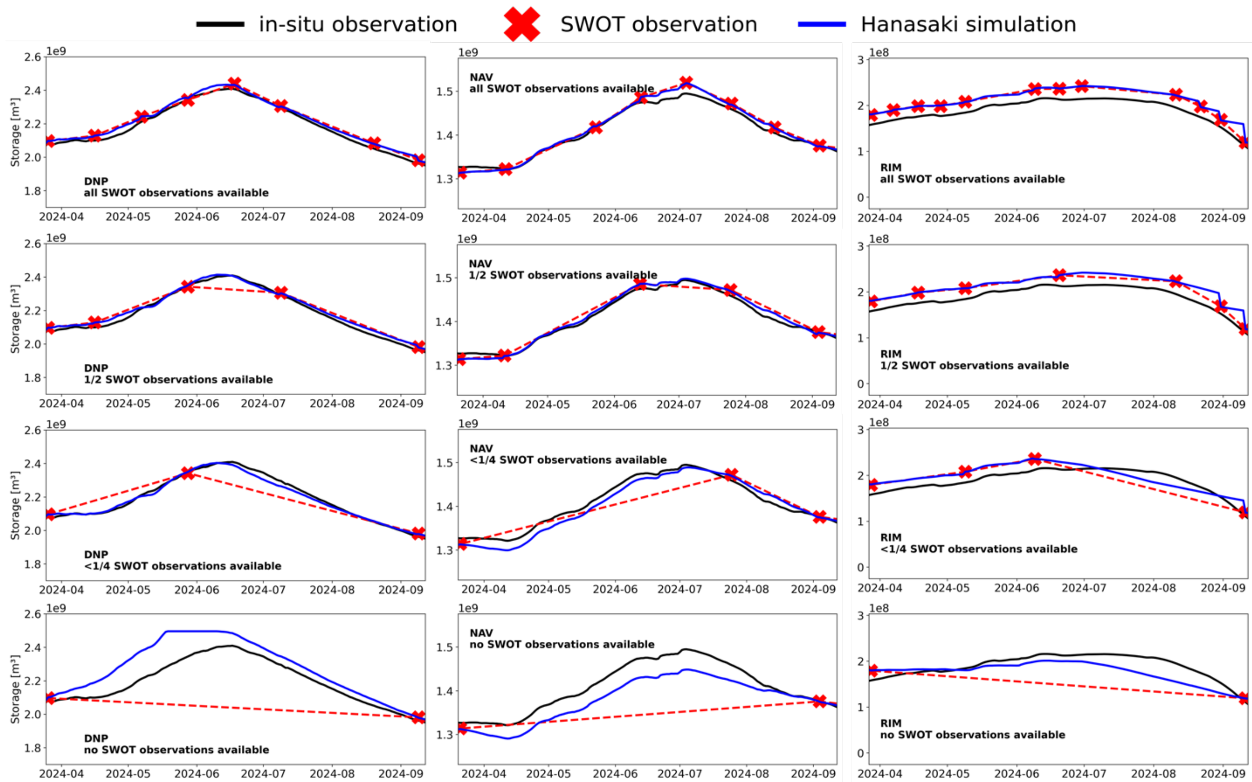


Figure 8. Time series of Hanasaki reservoir storage estimates when there are no (upper row), less than a quarter (second row), half (third row), and full (bottom row) SWOT observations available in DNP, NAV, and RIM.

In many cases, the SWOT observations are incomplete, especially for SWOT SA. For example, there are fewer than 50% SWOT SA observations actually available for DNP and NAV that meet the requirements given in section 2.2. Therefore, we explored use of the Hanasaki reservoir model to fill the temporal gaps in the SWOT observation records. We did so for three reservoirs, DNP in California, NAV in the USCR, and RIM in the CRB. From April 2024 to September 2024, there were more than 15 SWOT observations, of which 6, 6, and 10 were "good" and valid for DNP, NAV, and RIM, respectively. To test the performance of Hanasaki in filling the temporal gaps in SWOT observations, we assumed that, among the valid SWOT observations, there were full (scenario 1), half (scenario 2), less than a quarter (scenario 3), and no (scenario 4) valid SWOT observations between April 2024 to September 2024. For the Hanasaki-based reservoir storage simulations, we assumed the original (unscaled) reservoir inflows were known using available in-situ observations. However, we acknowledge this assumption will generally not hold true for most other reservoirs, where in-situ inflow data are typically unavailable. As described in section 2.5.3, when using the Hanasaki algorithm to fill temporal gaps in SWOT observations, we constrained the simulated storage series to match SWOT observations at the beginning and end of the period by scaling reservoir inflows.

For DNP in California (Figure 8 first column), between April 2024 to September 2024, there were 6 valid SWOT observations, so we assumed that there are 6, 3, 1, and 0 valid observations in our four scenarios. We found that at DNP, for scenarios 1, 2, and 3, the Hanasaki-based storage simulation maintains a high match with the in-situ storage observations, with KGE and R^2 both larger than 0.95. In scenario 3, Hanasaki simulated reservoir storage with the restriction of only one valid SWOT observation, while the simulation aligns well with in-situ observations. The largest relative error for scenario 3 was 1.75%, which benefits water managements even with very limited reservoir observations. When there were no SWOT observations (scenario 4), Hanasaki simulations mostly had a 20% – 30% positive bias compared to in-situ observations.

In the UCRB, similar to California, the Hanasaki algorithm with SWOT constraints performed well in scenario 1, 2, and 3, with increased bias in scenario 4 (Figure 8 second column). In scenario 1, the 1.60% positive bias is due primarily to overestimates of reservoir storage by SWOT because the Hanasaki simulation is forced to match SWOT observations when they are available. In that case, Hanasaki performs better in scenarios 2 and 3 than in scenario 1 for high reservoir volumes. With no SWOT observations during From April through September

2024, the Hanasaki simulation slightly underestimates reservoir storage by about 3% over the entire period, but this result is much better than that obtained by simply connecting SWOT storage estimates (as shown by the dashed red lines in Figure 8). For RIM in the CRB, the Hanasaki storage simulation performs worse than in the other basins, primarily due to the errors in SWOT estimates (Figure 8 third column). Even though SWOT WSEs have small errors relative to in-situ observations (Figure 3 third column), the overall positive biases in SWOT SA observations render SWOT-based reservoir storage estimates larger than in-situ. In SWOT observations, interpolation with the Hanasaki algorithm still provides continuous storage estimates with high accuracy. However, in cases where SWOT observations have large errors, Hanasaki estimates are degraded because we restricted the simulated storage series to match SWOT observation points. Furthermore, for the worst scenario when there are no SWOT observations during the five-month warm period, Hanasaki simulations nonetheless maintain an accuracy with absolute relative errors less than 20%.

Discussion

In general, we find that SWOT SA observations have much larger errors than WSE observations (sections 3.1 and 3.2). Nonetheless, SWOT-based reservoir storage estimate errors are mostly dominated by SWOT WSE errors. This in general benefits SWOT-based reservoir storage estimates. We also noticed that complete SWOT SA observations are very limited. We expect that future versions of the SWOT product (such as SWOT version D) will result in increases in the number of valid SWOT WSE observations, and in any event (in contrast to WSE) there are other options for estimating SA. For example, optical missions such as Landsat and Sentinel-2, or radar missions such as Sentinel-1, can provide complementary SA estimates to

SWOT. For most of our studied reservoirs, after screening out observations with “overlap” > 90% and with the acceptable observation quality (see Section 2.2), the remaining valid SA observations are usually less than 50% of the total possible. Compared to SWOT SA observations, SWOT WSE observations are more complete and with higher quality. Under such circumstances, for some reservoirs future studies might consider using a constant SA in lieu of the SWOT SA observations (or perhaps a robust smoother). This might help (1) avoid errors in SWOT-based storage estimates brought by large uncertainties in SWOT SA observations; (2) obtain a more temporally continuous reservoir storage estimate based on only WSE observations.

The reservoirs we selected are operated primarily for water supply. We made this choice primarily because their reservoir storage variations have distinct seasonal fluctuations that facilitate us to evaluate SWOT’s capability over a large range of water levels. Because the SWOT science phase data record is only about a year and a half, the water-supply reservoirs gave us the opportunity to study SWOT WSE and SA accuracy in comparison with in-situ observations over most of the physical storage range for the 12 reservoirs. Future studies should also consider reservoirs with other primary operating purposes, such as hydropower, flood control, and recreation. Another consideration leading to our focus on water supply reservoirs is that the reservoir operation algorithm we used to extend SWOT observations (Hanasaki) is primarily designed for water supply reservoirs. Other reservoir models, such as HYPE (Arheimer et al., 2020) could be leveraged to simulate the reservoir storage for hydropower reservoirs, and Döll (Döll et al., 2003) might be used for flood control reservoirs. While our analysis focuses on individually regulated reservoirs, future extensions could incorporate coordinated operation among cascaded reservoirs using multi-agent or cooperative reinforcement learning frameworks (e.g. Castelletti et al., 2010; Xu et al., 2021) to better represent human decision feedback and upstream–downstream dependencies.

An additional limitation is that we used observed reservoir inflows for the Hanasaki-based storage simulation. Such inflow observations are typically not available for reservoirs located outside of the WUS. In management applications where in situ observations aren't available, reservoir inflows could be simulated using hydrology models, estimated from SWOT-derived river discharge, providing spatially distributed inflow information in ungauged basins (Andreadis et al., 2025; Laipelt et al., 2025). We note, however, that the coarse temporal resolution and retrieval uncertainties in SWOT river slope and width (Durand et al., 2023b) currently limit the use of SWOT-derived discharge for reservoir operations. Another limitation is that we prescribed a baseline storage–WSE pair to convert storage changes into storage volumes. While this requires only a single pair per reservoir, it still typically depends on detailed in-situ historical data. Since such baseline pairs may not be available in most operational scenarios, future applications could leverage empirical relationships derived from remotely sensed data or generalized relationships calibrated from similar reservoirs to overcome this limitation.

Conclusions

Our results demonstrate that SWOT provides accurate measurements of reservoir water storage taken across twelve western U.S. water supply reservoirs with storage capacity ranging from 95 to 3019 MCM. Median WSE errors across all reservoirs and valid SWOT observations were under 20 cm and overall storage estimate errors below 10%. We also showed that model-based methods have the potential to fill temporal gaps in SWOT observations with relative errors < 20% for cases when as little as one-quarter of valid SWOT observations are available. These findings underscore the promise of advanced satellite interferometry for more reliable spatiotemporal monitoring of reservoir storage, thus paving the way for improved global water resource management.

More specifically, we find that:

1. SWOT can measure reservoir WSE with high accuracy in medium to large reservoirs (with capacity > 1000 MCM); however, measurement errors were large for the smaller reservoirs (with capacity < 500 MCM). The average MAE for SWOT WSE taken over all the reservoirs is < 10 cm, respectively, which is small compared to the seasonal WSE range for these reservoirs, which can be as large as hundreds of meters. Nonetheless, especially for the smaller reservoirs, there were large outliers, generally with 90th percentile errors > 100 cm.

2. SWOT-based SA measurements typically have much larger biases and uncertainties compared to SWOT WSE observations. However, errors in SWOT-derived reservoir storage estimates (our ultimate goal) are primarily determined by SWOT WSE errors. When accurate WSE measurements are available, SWOT-derived storage estimates align well with in-situ observations, with errors mostly less than 10%. Nonetheless, for smaller reservoirs that generally exhibit larger SA errors (e.g., TAY, RIM, HOW, and RIR), these errors may contribute to storage estimation errors.

3. In addition to evaluating SWOT's observational accuracy, our findings highlight the potential of reservoir operation algorithms (Hanasaki model) in extending the temporal continuity of storage records. Rather than solely depending on frequent SWOT passes, the

Hanasaki algorithm enables reconstruction of full reservoir storage time series with acceptable accuracy (<20% relative error), even under sparse (less than one quarter) or missing SWOT observations, making it especially valuable for long-term monitoring and operational forecasting in data-sparse conditions.

Overall, our study suggests the potential of SWOT for reservoir storage monitoring, particularly for medium to large sizes' reservoirs. We expect that algorithm advances and updated SWOT lake products (version D) will lead to more complete records, and reduce outliers, especially in SA estimates. This should lead to improvements for smaller reservoirs, which we suggest should be a focus of future research. Future research can also relax two of our key assumptions; one being the constraining of storage variations to match an observed (baseline) WSE-storage pair, and the other the use of observed reservoir inflows in connection with the Hanasaki storage algorithm (an obvious alternative being use of global or regional hydrologic model output).

Data Availability Statement

Reservoir-related data, including in-situ and SWOT observations of water surface elevation (WSE), surface area (SA), and derived reservoir storage, are archived in a public repository (https://figshare.com/s/eae6081b9b3a777f2cf6, Ma et al., 2025).

References

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., et al. (2010). The ICESat-2 Laser Altimetry Mission. *Proceedings of the IEEE*, *98*(5), 735–751. https://doi.org/10.1109/JPROC.2009.2034765

Adam, J. C., Haddeland, I., Su, F., & Lettenmaier, D. P. (2007). Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers. *Journal of Geophysical Research: Atmospheres*, *112*(D24). https://doi.org/10.1029/2007JD008525

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*(2). https://doi.org/10.1029/2006RG000197

Andreadis, K. M., Coss, S. P., Durand, M., Gleason, C. J., Simmons, T. T., Tebaldi, N., et al. (2025). A First Look at River Discharge Estimation From SWOT Satellite Observations. *Geophysical Research Letters*, *52*(9), e2024GL114185. https://doi.org/10.1029/2024GL114185

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2020). Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, *24*(2), 535–559. https://doi.org/10.5194/hess-24-535-2020

Avisse, N., Tilmant, A., Müller, M. F., & Zhang, H. (2017). Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas. *Hydrology and Earth System Sciences*, *21*(12), 6445–6459. https://doi.org/10.5194/hess-21-6445-2017

Benson, R. D. (2018). Reviewing reservoir operations in the North American West: an opportunity for adaptation. *Regional Environmental Change*, *18*(6), 1633–1643. https://doi.org/10.1007/s10113-018-1330-x

Berry, P. a. M., Garlick, J. D., Freeman, J. A., & Mathers, E. L. (2005). Global inland water monitoring from multi-mission altimetry. *Geophysical Research Letters*, *32*(16). https://doi.org/10.1029/2005GL022814

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., et al. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, *47*(3). https://doi.org/10.1029/2009WR008929

Birkett, C. M. (1994). Radar altimetry: A new concept in monitoring lake level changes. *Eos, Transactions American Geophysical Union*, *75*(24), 273–275. https://doi.org/10.1029/94EO00944

Birkett, Charon M., O'Brien, K., Kinsey, S., Ricko, M., & Li, Y. (2022). Enhancement of a global lake and reservoir database to aid climate studies and resource monitoring utilizing satellite radar altimetry. *Journal of Great Lakes Research*, *48*(1), 37–51. https://doi.org/10.1016/j.jglr.2021.11.013

Birol, F., Bignalet-Cazalet, F., Cancet, M., Daguze, J.-A., Fkaier, W., Fouchet, E., et al. (2025). Understanding uncertainties in the satellite altimeter measurement of coastal sea level: insights from a round-robin analysis. *Ocean Science*, *21*(1), 133–150. https://doi.org/10.5194/os-21-133-2025

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., et al. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669–690. https://doi.org/10.5194/hess-23-669-2019

Castelletti, A., Galelli, S., Restelli, M., & Soncini-Sessa, R. (2010). Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research*, *46*(9). https://doi.org/10.1029/2009WR008898

Crétaux, J.-F., & Birkett, C. (2006). Lake studies from satellite radar altimetry. *Comptes Rendus Geoscience*, *338*(14), 1098–1112. https://doi.org/10.1016/j.crte.2006.08.002

Das, P., & Hossain, F. (2025). Multi-Satellite Tracking of Surface Water Storage Change in the Era of Surface Water and Ocean Topography (SWOT) Satellite Mission. *Earth and Space Science*, *12*(6), e2024EA004178. https://doi.org/10.1029/2024EA004178

Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, *13*(12), 2413–2432. https://doi.org/10.5194/hess-13-2413-2009

Döll, Petra, Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, *270*(1), 105–134. https://doi.org/10.1016/S0022-1694(02)00283-4

Domeneghetti, A., Schumann, G. J.-P., Frasson, R. P. M., Wei, R., Pavelsky, T. M., Castellarin, A., et al. (2018). Characterizing water surface elevation under different flow conditions for the upcoming SWOT mission. *Journal of Hydrology*, *561*, 848–861. https://doi.org/10.1016/j.jhydrol.2018.04.046

Durand, M., Chen, C., de Moraes Frasson, R. P., Pavelsky, T. M., Williams, B., Yang, X., & Fore, A. (2020). How will radar layover impact SWOT measurements of water surface elevation and slope, and estimates of river discharge? *Remote Sensing of Environment*, *247*, 111883. https://doi.org/10.1016/j.rse.2020.111883

Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., et al. (2023a). A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission. *Water Resources Research*, *59*(4), e2021WR031614. https://doi.org/10.1029/2021WR031614

Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., et al. (2023b). A Framework for Estimating Global River Discharge From the Surface Water and Ocean Topography Satellite Mission. *Water Resources Research*, *59*(4), e2021WR031614. https://doi.org/10.1029/2021WR031614

Frappart, F., Tong Minh, D. H., Baghdadi, N., Crétaux, J.-F., Fayad, I., & Bergé-Nguyen, M. (2024). Improving mean water lake surface elevation estimates using dense lidar measurements from the GEDI satellite mission. *Remote Sensing Applications: Society and Environment*, *35*, 101213. https://doi.org/10.1016/j.rsase.2024.101213

Gao, H. (2015a). Satellite remote sensing of large lakes and reservoirs: from elevation and area to storage. *WIREs Water*, *2*(2), 147–157. https://doi.org/10.1002/wat2.1065

Gao, H. (2015b). Satellite remote sensing of large lakes and reservoirs: from elevation and area to storage. *WIREs Water*, *2*(2), 147–157. https://doi.org/10.1002/wat2.1065

Gao, H. (2015c). Satellite remote sensing of large lakes and reservoirs: from elevation and area to storage. *WIREs Water*, *2*(2), 147–157. https://doi.org/10.1002/wat2.1065

Gao, H., Birkett, C., & Lettenmaier, D. P. (2012a). Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, *48*(9). https://doi.org/10.1029/2012WR012063

Gao, H., Birkett, C., & Lettenmaier, D. P. (2012b). Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, *48*(9). https://doi.org/10.1029/2012WR012063

Gharari, S., Vanderkelen, I., Tefs, A., Mizukami, N., Kluzek, E., Stadnyk, T., et al. (2024). A Flexible Framework for Simulating the Water Balance of Lakes and Reservoirs From Local to Global Scales: mizuRoute-Lake. *Water Resources Research*, *60*(5), e2022WR032400. https://doi.org/10.1029/2022WR032400

Guan, X., Jiang, P., Meng, Y., Qin, H., & Lv, H. (2020). Study on Production, Domestic and Ecological Benefits of Reservoir Water Supply Based on Emergy Analysis. *Processes*, *8*(11), 1435. https://doi.org/10.3390/pr8111435

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hamoudzadeh, A., Ravanelli, R., & Crespi, M. (2024). SWOT Level 2 Lake Single-Pass Product: The L2_HR_LakeSP Data Preliminary Analysis for Water Level Monitoring. *Remote Sensing*, *16*(7), 1244. https://doi.org/10.3390/rs16071244

Hanazaki, R., Yamazaki, D., & Yoshimura, K. (2022). Development of a Reservoir Flood Control Scheme for Global Flood Models. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002944. https://doi.org/10.1029/2021MS002944

Heffernan, S., & Strimbu, B. M. (2021). Estimation of Surface Canopy Water in Pacific Northwest Forests by Fusing Radar, Lidar, and Meteorological Data. *Forests*, *12*(3), 339. https://doi.org/10.3390/f12030339

Hosseini-Moghari, S.-M., & Döll, P. (2025). The value of observed reservoir storage anomalies for improving the simulation of reservoir dynamics in large-scale hydrological models. *Hydrology and Earth System Sciences*, *29*(17), 4073–4092. https://doi.org/10.5194/hess-29-4073-2025

Jahandideh-Tehrani, M., Bozorg Haddad, O., & Loáiciga, H. A. (2015). Hydropower Reservoir Management Under Climate Change: The Karoon Reservoir System. *Water Resources Management*, *29*(3), 749–770. https://doi.org/10.1007/s11269-014-0840-7

James, L. D. (1970). Economic Optimization and Reservoir Recreation. *Journal of Leisure Research*, *2*(1), 16–29. https://doi.org/10.1080/00222216.1970.11969980

Kica, S., Pavelsky, T. M., Fayne, J. V., & Williams, B. A. (2025). SWOT Water Surface Elevation in Herbaceous Wetlands of Florida's Everglades. *Geophysical Research Letters*, *52*(9), e2025GL114956. https://doi.org/10.1029/2025GL114956

Laipelt, L., de Paiva, R. C. D., Fan, F. M., Collischonn, W., Papa, F., & Ruhoff, A. (2025). SWOT Reveals How the 2024 Disastrous Flood in South Brazil Was Intensified by Increased Water Slope and Wind Forcing. *Geophysical Research Letters*, *52*(2), e2024GL111287. https://doi.org/10.1029/2024GL111287

Lu, S., Dai, W., Tang, Y., & Guo, M. (2020). A review of the impact of hydropower reservoirs on global climate change. *Science of The Total Environment*, *711*, 134996. https://doi.org/10.1016/j.scitotenv.2019.134996

Patidar, G., Indu, J., & Karmakar, S. (2025). Performance Assessment of Surface Water and Ocean Topography (SWOT) Mission for WSE Measurement Across India. *Geophysical Research Letters*, *52*(10), e2025GL115804. https://doi.org/10.1029/2025GL115804

Qian, B., Zhang, D., Wang, J., Huang, F., & Wu, Y. (2016). Impacts of reservoirs on the streamflow and sediment load of the Hanjiang River, China. *Environmental Monitoring and Assessment*, *188*(11), 646. https://doi.org/10.1007/s10661-016-5652-1

Ren, M., He, X., Kan, G., Wang, F., Zhang, H., Li, H., et al. (2017). A Comparison of Flood Control Standards for Reservoir Engineering for Different Countries. *Water*, *9*(3), 152. https://doi.org/10.3390/w9030152

Sadki, M., Munier, S., Boone, A., & Ricci, S. (2023). Implementation and sensitivity analysis of the Dam-Reservoir OPeration model (DROP v1.0) over Spain. *Geoscientific Model Development*, *16*(2), 427–448. https://doi.org/10.5194/gmd-16-427-2023

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., & DiMarzio, J. P. (2005). Overview of the ICESat Mission. *Geophysical Research Letters*, *32*(21). https://doi.org/10.1029/2005GL024009

Solander, K. C., Reager, J. T., & Famiglietti, J. S. (2016). How well will the Surface Water and Ocean Topography (SWOT) mission observe global reservoirs? *Water Resources Research*, *52*(3), 2123–2140. https://doi.org/10.1002/2015WR017952

Sulistioadi, Y. B., Tseng, K.-H., Shum, C. K., Hidayat, H., Sumaryono, M., Suhardiman, A., et al. (2015). Satellite radar altimetry for monitoring small rivers and lakes in Indonesia. *Hydrology and Earth System Sciences*, *19*(1), 341–359. https://doi.org/10.5194/hess-19-341-2015

Tang, Q., Gao, H., Lu, H., & Lettenmaier, D. P. (2009). Remote sensing: hydrology. *Progress in Physical Geography: Earth and Environment*, *33*(4), 490–509. https://doi.org/10.1177/0309133309346650

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., et al. (2022a). Evaluating a reservoir parametrization in the vector-based global routing model mizuRoute (v2.0.1) for Earth system model coupling. *Geoscientific Model Development*, *15*(10), 4163–4192. https://doi.org/10.5194/gmd-15-4163-2022

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., et al. (2022b). Evaluating a reservoir parametrization in the vector-based global routing model mizuRoute (v2.0.1) for Earth system model coupling. *Geoscientific Model Development*, *15*(10), 4163–4192. https://doi.org/10.5194/gmd-15-4163-2022

Vicente-Serrano, S. M., Zabalza-Martínez, J., Borràs, G., López-Moreno, J. I., Pla, E., Pascual, D., et al. (2017). Effect of reservoirs on streamflow and river regimes in a heavily regulated river basin of Northeast Spain. *CATENA*, *149*, 727–741. https://doi.org/10.1016/j.catena.2016.03.042

Wang, J., Pottier, C., Cazals, C., Battude, M., Sheng, Y., Song, C., et al. (2025). The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake mask and operational auxiliaries. Retrieved from https://www.authorea.com/doi/full/10.22541/au.170258987.72387777?commit=681b4cb8c7423197eb330cacf5a11705e2c93979

Xu, W., Meng, F., Guo, W., Li, X., & Fu, G. (2021). Deep Reinforcement Learning for Optimal Hydropower Reservoir Operation. *Journal of Water Resources Planning and Management*, *147*(8), 04021045. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001409

Yu, L., Zhang, H., Gong, W., & Ma, X. (2024). Validation of Mainland Water Level Elevation Products From SWOT Satellite. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 13494–13505. https://doi.org/10.1109/JSTARS.2024.3435363

Chapter 4. SWOT-Initialized Reservoir Operation Forecasting: Validation in Data-Rich and Application to Data-Limited Regions

This chapter is currently under review by *Hydrology and Earth System Sciences*.

Abstract

Challenges in reservoir operation are posed by the fact that reservoir release trajectories depend on both uncertain inflow forecasts and (in cases where in situ observations are not available) imperfect knowledge of the initial reservoir state. The recently launched (science data available since late 2023) Surface Water and Ocean Topography (SWOT) satellite provides global-scale reservoir water level and surface extent observations, creating new opportunities to constrain reservoir storage states for forecast initialization. Here we develop a SWOT-initialized reservoir operation forecasting framework and evaluate it at two data-rich California reservoirs and at a data-limited reservoir in Ethiopia. The operations framework combines SWOT-derived reservoir storage, monthly inflow forecasts, a reservoir operation model, and a downstream agricultural water demand-satisfaction assessment. Results from California reservoirs show that SWOT-derived storage provides useful initial conditions for seasonal reservoir operation forecasts (similar to those available based on in situ data), with water demand satisfaction approaching that obtained using in-situ storage initialization. The resulting storage and release forecasts capture the major seasonal dynamics. Although Ensemble Streamflow Prediction (ESP) (developed using historic observed reservoir inflows) generally provides more accurate inflow and reservoir-operation forecasts, a global reservoir inflow product (GloFAS) produced without local observations yields water demand satisfaction patterns that are broadly consistent with the in situ results. Application of the framework to the African reservoir (where essentially no in situ

observations are available) demonstrates that, for the two years of SWOT data, nominal irrigation water demands are generally satisfied.

4.1 Introduction

Reservoirs play a critical role in water-resources management by regulating streamflow, which in turn assures the reliability of municipal and agricultural water supply, mitigates floods, and supports in-stream environments (Ahmad et al., 2014; Yan et al., 2024). In the western United States (WUS), reservoirs are particularly important because water availability is strongly seasonal, interannual variability is large, and droughts and atmospheric-river events can create rapid transitions between water scarcity and floods (Dettinger and Anderson, 2015; Spinti et al., 2023). Reservoir operators therefore rely on forecasts of future inflow, storage, and release over time scales ranging from days to seasons (e.g., Kim et al., 2022; McGuire et al., 2006; Özdoğan-Sarıkoç et al., 2023; Sushanth et al., 2023). Such forecasts can inform reservoir release planning, drought preparedness, flood control, and water allocation decisions (Delaney et al., 2020; Zanutto et al., 2025), but their usefulness depends on the quality of reservoir inflow and storage forecasts.

Reservoir storage forecast errors arise from two primary sources. The first is uncertainty in future inflow, which reflects uncertainty in meteorological forcing, runoff generation, snowmelt, and upstream flow conditions (Rahmanifard and Gates, 2024). The second (important in cases where in situ data are not available) is uncertainty in the initial reservoir storage state, which constrains the subsequent storage and release trajectory (Tian et al., 2022). Initial-storage errors can dominate storage forecasts at short lead times, whereas accumulated inflow uncertainty dominates as forecast lead time increases (Yang et al., 2021; Zhao et al., 2011). A useful reservoir forecasting system must therefore distinguish the contribution of initialization error from the contribution of inflow forecast error and evaluate how both propagate into operationally relevant release forecasts.

In the developed world, reservoir stage (water level) is usually measured with in situ stage recorders. In combination with bathymetric surveys, the stage information is used to produce stage-storage relationships, so real-time or near real-time stage measurements can be converted to storage. Under these conditions, the largest source of storage error is bathymetric error, which generally increase over time due to reservoir sedimentation. Alternatively (and of special importance in transboundary rivers where in situ stage information may not be available to downstream water users) satellite remote sensing provides an opportunity to estimate reservoir storage (e.g., Sentinel-1: Torres et al., 2012; Sentinel-2: Drusch et al., 2012; SWOT: Biancamaria et al., 2016). The recently launched Surface Water and Ocean Topography (SWOT) mission (data available since mid-2023) represents a major advance in satellite hydrology by providing, for the first time, spatially extensive and simultaneous observations of reservoir water-surface elevation (WSE) and surface extent, enabling consistent estimation of storage and storage change for lakes and reservoirs in data-sparse regions (Fu et al., 2024). As we have shown in previous work (Ma et al., 2026) SWOT observations can produce reservoir storage estimates with useful accuracy for medium and large reservoirs.

Existing studies have advanced satellite-based reservoir storage estimation (e.g., Gao et al., 2012; Ma et al., 2026), streamflow forecasting (e.g., Maddu et al., 2022; Yang et al., 2019), and state updating in reservoir systems (e.g., Iglesias et al., 2013; Zhang et al., 2021). Nevertheless, most satellite-reservoir studies emphasize historical monitoring or storage reconstruction rather than reservoir-operation forecasts. In addition, storage accuracy alone does not establish whether a forecast is useful for water management: inflow (forecast) errors strongly affect reservoir storage forecasts, and reservoir releases (which determine whether downstream water demands are likely to be met) are strongly affected by reservoir storage errors that in turn depend on inflow forecasts. A remaining need is therefore to connect satellite-derived storage initialization with seasonal inflow forecasts, reservoir operation modeling, and demand-oriented evaluation in a single transferable workflow.

Here, we address the above need by developing a SWOT-initialized reservoir operation forecasting framework, which we initially apply to two data-rich WUS reservoir systems, and then to a data-limited reservoir in the Ethiopian Highlands. In the data-rich WUS cases, monthly forecasts are made on regular forecast dates (the 10th of each month) rather than only on SWOT overpass dates. For each forecast date, we propagate the most recent SWOT-derived storage to each forecast date with observed inflows using the Hanasaki reservoir operation model (Hanasaki et al., 2006). The resulting storage serves as the SWOT-based initial condition. Future inflows come from two seasonal forecasting approaches: locally derived Ensemble Streamflow Prediction (ESP) and the globally available Global Flood Awareness System (GloFAS) seasonal forecasts. The Hanasaki model converts these inflow scenarios into storage and release forecasts, which we evaluate relative to observations and downstream water-use demands.

The WUS reservoirs provide a controlled evaluation environment given that in-situ storage, inflow, and release observations are available for comparison. Our broader purpose, however, is to transfer the methods to data-limited settings in which satellite storage and global inflow forecasts may be available even when continuous (in situ) reservoir observations are not. This study is organized around one overarching question: can SWOT-derived reservoir storage enable reservoir operation forecasting that informs downstream water-management decisions, especially in regions with limited in-situ observations? We address this question through four linked components: the accuracy of SWOT-based initialization, the performance of seasonal inflow forecasts, the propagation of those uncertainties into storage and reservoir releases, and the ability of forecasted releases to meet agricultural water demands (which is the primary water use in all three case studies).

4.2 Study Areas

4.2.1 Data-Rich Validation Reservoirs: BUL and DNP

We first evaluate our water management framework at two well-observed reservoirs in the WUS: New Bullards Bar Reservoir (BUL; Figure 4.1a) and Don Pedro Reservoir (DNP; Figure 4.1b). Both reservoirs drain the western slopes of California's Sierra Nevada (see Figure 4.1) and have sufficiently complete in-situ storage, inflow, and reservoir release records to evaluate our reservoir storage and release forecast framework. The contrasting reservoir storage capacities and operating settings across the two sites provide complementary validation cases for assessing the forecasting framework.

BUL is located on the North Yuba River and has a maximum storage capacity of approximately 1,192 million m^3 (Mm^3). DNP is located on the Tuolumne River and is substantially larger, with a maximum storage capacity of approximately 2,504 Mm^3 (Ma et al., 2026). During the SWOT analysis period, both reservoirs reached more than 95% of their reported maximum storage capacities and exhibited pronounced seasonal filling and drawdown. These conditions provide observations spanning reservoir filling, near-full storage, and drawdown and therefore support evaluation of both initialization error and inflow-driven forecast error. We calculated reservoir storage-to-inflow ratios from the in situ maximum storage capacity and mean annual inflow for each reservoir. Although the ratios vary among reservoirs, all are well below 1.0, indicating limited interannual storage memory and suggesting that reservoir operations are dominated by within-year rather than over-year regulation.

Because extensive in situ observations are available for BUL and DNP, they are used to evaluate the SWOT-informed reservoir forecasting framework before its transfer to the data-limited setting. The BUL and DNP in-situ records allow the SWOT-based initial storage state to be compared with observed storage on each forecast issue date and permit evaluation of forecasted inflow, storage, and release. This controlled validation is necessary before applying the framework where equivalent observations are sparse.

4.2.2 Data-Limited Application: Tendaho Reservoir

Tendaho Irrigation Reservoir in the Lower Awash Basin, Ethiopia (Figure 4.1c), is used to demonstrate transfer of the validated workflow to a data-limited setting. Tendaho is an irrigation-focused reservoir for which continuous storage observations, detailed release records, and operating-rule information are more limited than for the WUS reservoirs. At the same time, SWOT observations, historical inflow information, and project-scale irrigation-demand estimates are available. Tendaho is therefore treated as a transfer application rather than as an equally data-rich validation experiment.

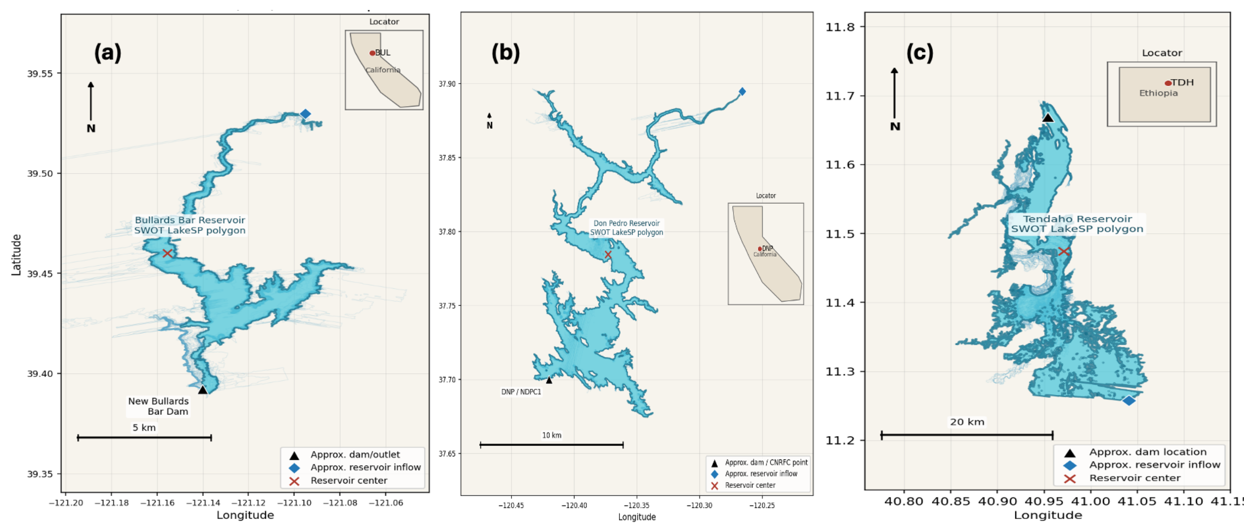


Figure 4.1. Study reservoirs and associated SWOT LakeSP (lake single pass) polygons: (a) New Bullards Bar Reservoir, California, USA; (b) Don Pedro Reservoir, California, USA; and (c) Tendaho Reservoir, Ethiopia. Symbols indicate the approximate dam/outlet location, reservoir inflow location, and reservoir center. Insets show the regional location of each reservoir.

4.3 Data

Our analysis combines four categories of data: SWOT observations of reservoir water-surface elevation and area; in-situ reservoir observations (BUL and DNP only) used for validation; seasonal reservoir inflow forecasts; and reservoir attributes and agricultural water-demand information used for reservoir modeling and demand-satisfaction assessment. The western U.S. reservoirs BUL and DNP provide relatively complete observational benchmarks, whereas the Tendaho application relies primarily on SWOT observations, globally available reservoir inflow forecasts, and limited historical and project-level information.

4.3.1 SWOT LakeSP Observations

The SWOT Level 2 Lake Single-Pass Vector Data Product, Version D (LakeSP) (NASA/JPL PO.DAAC, 2026) provides reservoir water-surface elevation (WSE) and surface area. We derived reservoir storage by adapting the reference-based framework of Ma et al. (2026). For each reservoir, we selected the maximum quality-controlled WSE observed during the SWOT science phase (August 2023–present) as the reference WSE and assumed that it represented the full-capacity condition. We used the reported design maximum storage capacity (e.g., from the Global Reservoir and Dam Database (GRanD)) as the corresponding reference storage, which we assumed was equal to the maximum inferred (from SWOT observations) storage during the period of SWOT observations. We then estimated storage for each quality-controlled SWOT observation by calculating the storage change relative to this reference point from the difference between the observed and reference WSEs, multiplied by the corresponding SWOT surface area, and adding this change to the reference storage. Unlike the approach in Ma et al. (2026), this method does not require a paired in situ WSE–storage observation and therefore facilitates application in data-limited regions where reservoir observations are sparse or unavailable.

Following Ma et al. (2026), we excluded SWOT observations with insufficient reservoir spatial coverage or degraded quality. For each retained overpass, we estimated reservoir storage relative to a reservoir-specific reference WSE–storage pair using the quality-controlled SWOT-observed WSE and the corresponding SWOT-observed surface area from the SWOT LakeSP product. We used the observed surface area together with the change in WSE to estimate the storage change from the reference condition (defined by pairing the maximum quality-controlled SWOT WSE observed during the science phase with the reported maximum storage capacity of each reservoir). No SWOT-based storage estimate was produced for overpasses that failed the quality-control criteria. We then used the resulting irregularly spaced (in time) storage estimates as satellite-derived initialization anchors, as described in Section 4.4.3.

4.3.2 In Situ Reservoir Observations

In situ observations provide the benchmark for our WUS validation experiments. We obtained daily or near-daily WSE, storage, inflow, and release records from the Yuba Water Agency for New Bullards Bar Reservoir (BUL) and from the U.S. Bureau of Reclamation (USBR) for Don Pedro Reservoir (DNP) (Ma et al., 2026). The reference storage values used to evaluate SWOT-derived storage are distinct from the daily in situ storage records used to construct the observed-storage initialization benchmark and to evaluate storage and release simulations from the Hanasaki reservoir model. We also used the in situ inflow records to evaluate the seasonal inflow forecasts.

We expressed all reservoir storage values as absolute storage so that SWOT-derived, propagated, simulated, and observed storage states (see Section 4.3.1) can be compared directly. In so doing, the underlying assumption is that each reservoir approached its reported maximum storage capacity during the SWOT observation period. This assumption appears reasonable for the two California reservoirs because water years 2024 and 2025 were relatively wet and both reservoirs reached near-capacity conditions during this period.

4.3.3 Seasonal Inflow Forecasts

We used two seasonal inflow forecast data sets for the WUS reservoirs. Ensemble Streamflow Prediction (ESP; Day, 1985) essentially resamples hydrometeorological conditions from the historic past to produce streamflow ensembles that are initialized with hydrologic storage (soil moisture and snow water equivalent) on each forecast date. In our case, we used the Noah-MP (Niu et al., 2011) hydrologic model, which we calibrated using historical hydrometeorological data including observed reservoir inflows for BUL and DNP to produce ESP forecasts. In addition, we used the Global Flood Awareness System (GloFAS) ensemble forecasts produced by ECMWF. GloFAS provides consistent seasonal streamflow forecasts initialized at regular monthly intervals (Alfieri et al., 2013; Emerton et al., 2018). We bias corrected the GloFAS inflow forecasts using correction relationships derived from historical GloFAS simulations and corresponding in situ inflow observations (see Figures B1 and B2 in Section S1).

For Tendaho, the only inflow forecasts available to us were from GloFAS (ESP was not feasible as we lacked the historical hydrometeorological data that it requires). We corrected systematic bias using historical in-situ inflow observations from the period 1970-2000, and we applied the resulting corrections to the seasonal GloFAS forecasts for the SWOT period (no streamflow observations were available during the SWOT period).

4.3.4 Reservoir Attributes and Agricultural Water-Demand Data

Reservoir-specific inputs include maximum storage capacity, elevation-storage information, mean annual inflow, and the demand characteristics required by the reservoir operation model. We fixed these attributes during the paired initialization and inflow-forecast experiments so that differences in forecast performance can be attributed to the initial storage source and inflow forecast source.

For BUL, we estimated mean annual downstream agricultural demand from California Department of Water Resources (DWR) county-level agricultural applied water estimates for Yuba County for water years 1998–2020 (California Department of Water Resources, 2024 a,b,c,d). For DNP, we estimated downstream agricultural demand from DWR WUEData Aggregated Farm-Gate Delivery records for Turlock Irrigation District (TID) and Modesto Irrigation District (MID) (California Department of Water Resources, n.d.; https://wuedata.water.ca.gov/). MID reports monthly farm-gate deliveries, whereas TID reports bimonthly deliveries; the bimonthly TID values were split evenly across the two months. Historical demand profiles were calculated from 2012–2023 TID+MID farm-gate deliveries.

We then allocated the annual totals across months using Spatial CIMIS reference evapotranspiration fractions (California Department of Water Resources, n.d.), calculated as each month's reference evapotranspiration divided by the annual total. This approach assumes that the seasonal distribution of irrigation demand follows atmospheric evaporative demand. We estimated a month-specific agricultural release fraction as the ratio of ET-distributed irrigation demand to historical mean monthly release. We multiplied forecasted total release by this fraction to obtain a proxy for agricultural water supply. For Tendaho, we distributed annual project-scale irrigation demand (Ministry of Irrigation and Lowlands, n.d.) seasonally using normalized monthly net irrigation requirement values reported by Shonka and Gullie (2017).

4.4 Methods

4.4.1 Forecasting Framework and Experiment Design

The forecasting framework consists of four linked steps (Figure 4.2). First, quality-controlled SWOT WSE and surface area observations are converted to reservoir storage. Second, the most recent SWOT-derived storage is propagated to each regular forecast issue date (10th of every month) and used to initialize the Hanasaki reservoir operation model (Hanasaki et al., 2006). Third, ESP and/or GloFAS seasonal inflow forecasts are used to simulate future storage and release. Fourth, forecasts are evaluated against available observations, and forecasted release is compared with agricultural demand using a demand-satisfaction ratio. For BUL and DNP, paired experiments combined either observed or SWOT-based storage initialization with either ESP or GloFAS inflow. Tendaho used the same workflow but with only SWOT initialization, bias-corrected GloFAS inflow, and literature-based irrigation demand.

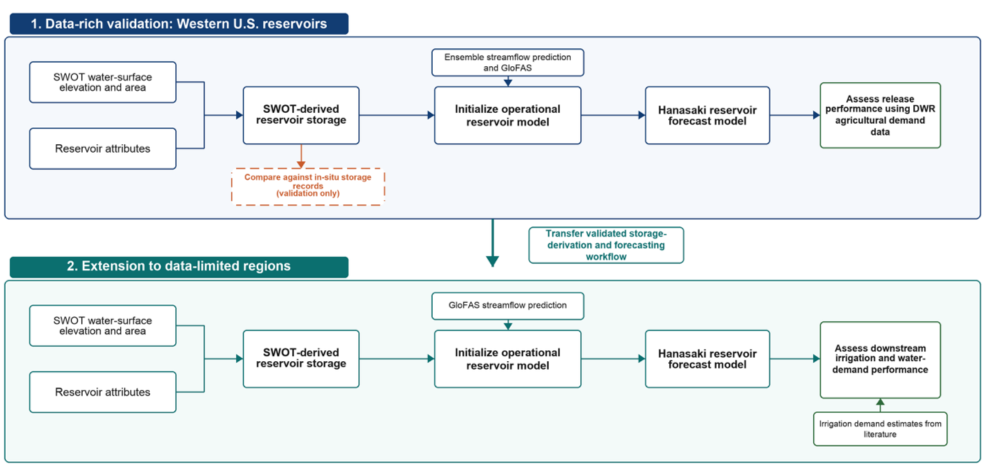


Figure 4.2. Conceptual framework for validating and transferring a SWOT-informed reservoir forecasting workflow from data-rich to data-limited regions.

4.4.2 Seasonal Reservoir Inflow Forecasts

As noted above, we generated reservoir inflow forecasts at BUL and DNP using both ESP and GloFAS, and at Tendaho using GloFAS seasonal forecasts. ESP provides a locally configured ensemble of future inflow trajectories by combining the modeled hydrologic state at each forecast issue date with meteorological forcing sequences sampled from historical years. GloFAS provides globally consistent seasonal streamflow forecasts initialized at regular monthly intervals, enabling reservoir inflow forecasting across regions worldwide, including basins where locally configured forecasting systems are unavailable.

We generated ESP forecasts at BUL and DNP using the Noah-MP land-surface model (Niu et al., 2011) coupled with the WRF-Hydro routing framework (Gochis et al., 2020). The model configuration follows the western U.S. implementation described by Su et al. (2025) and Ma et al. (2026), including the land-surface parameterization, routing-domain configuration, spatial resolution, and basin-specific setup used to simulate reservoir inflow. We extracted simulated streamflow at the grid cells corresponding to the reservoir inflow locations and converted from discharge to inflow volume for use in the reservoir operation simulations.

We obtained historical meteorological forcings for Noah-MP from the 1-km continental United States forcing dataset developed by Pan et al. (2025). The forcing variables include precipitation, near-surface air temperature, surface pressure, downward shortwave radiation, downward longwave radiation, wind speed, and specific humidity. The historical forcing record spans 1980–present and is available at a temporal resolution of 1 h. We used these meteorological forcings to drive Noah-MP and WRF-Hydro for both the retrospective model simulations for water years 2020–2022 (for spin-up) and for the generation of ESP ensemble members during the forecast periods in 2024 and 2025 (see Section 4.3.3).

For each monthly forecast issue date, we first initialized the hydrologic model using the modeled land-surface and routing states corresponding to that date. These states include soil moisture, snow water equivalent, groundwater or subsurface storage, channel states, and other prognostic variables represented by the Noah-MP–WRF-Hydro system. We then represented future meteorological uncertainty by replacing the unknown future forcings with meteorological sequences drawn from individual historical years in the retrospective period 1980–2023. Each historical forcing year is applied from the same forecast-date hydrologic state, producing one possible future inflow trajectory. Therefore, for a given forecast cycle, all 44 ensemble members share the same initial hydrologic conditions but differ in their subsequent meteorological forcing.

We used historical forcing sequences from 1980 to 2023 to generate 44 ESP ensemble members for each forecast issue date. We initialized forecasts on the 10th of each month and continued to the end of the water year (September 30). We aggregated the model outputs from the native 3-hourly to monthly inflow volumes. We summarized each ESP ensemble using the ensemble mean, the interquartile range, and the 2^{nd}–98^{th} percentile range to represent forecast uncertainty. We then used the same ESP inflow ensemble to force the reservoir operation model for the alternative storage-initialization experiments. Specifically, we initialized reservoir simulations using either observed in-situ storage or SWOT-derived storage, while the inflow ensemble and reservoir-model configuration were held constant. This design allows the effect of initial-storage uncertainty to be evaluated separately from uncertainty in the future inflow forecast.

We obtained GloFAS seasonal streamflow forecasts from version 4.0 of the GloFAS seasonal forecasting system (Joint Research Centre and Copernicus Emergency Management Service, 2020). As for ESP, we initialized forecasts on the 10^{th} of each month. The GloFAS ensembles each contain 51 ensemble members and extend to a maximum lead time of ~7 months. We extracted streamflow from the GloFAS grid cell located at or immediately upstream of each reservoir inflow location. Where necessary, we converted discharge to reservoir inflow volume and aggregated to the same temporal resolution used for the ESP forecasts.

For BUL and DNP, we first compared ESP and GloFAS forecasts with observed reservoir inflow. We then passed the two inflow products through the same reservoir operation model (Hanasaki) using identical reservoir parameters and storage-initialization alternatives. Because we held the forecast issue dates, reservoir-model configuration, and initial-condition experiments constant, differences in simulated storage, release, and demand-satisfaction performance can be attributed primarily to differences in inflow forecast quality.

For Tendaho Reservoir, an equivalent locally configured Noah-MP–WRF-Hydro ESP system was not available. GloFAS therefore provides the sole seasonal reservoir inflow forecast. We corrected systematic bias in GloFAS inflow using available historical in-situ inflow observations from 1980 to 2000. We derived the bias-correction method from monthly means over the historical overlap period and then applied it to the seasonal GloFAS forecasts. We then used the bias-corrected GloFAS ensemble to drive the same reservoir operation model applied in the western U.S. experiments. This configuration allowed the forecasting workflow to be transferred to a data-limited setting using SWOT-derived storage, a globally available seasonal inflow product, and limited historical information.

4.4.3 SWOT-Based Reservoir-Storage Initialization

The key methodological step in our forecast system is construction of a SWOT-based storage initial condition at each regular monthly forecast issue date, T0, which we took as the 10th of each month. Because SWOT overpass dates generally do not coincide with T0, we identified the latest valid SWOT-derived storage observation before T0 and propagated forward with the reservoir operation model using simulated ESP and GloFAS inflows over the intervening period. The resulting state is referred to as the SWOT-initialized storage at T0.

We compared two initial-condition experiments for the WUS reservoirs. The benchmark experiment uses observed in-situ storage at T0, whereas the satellite-based experiment uses the propagated SWOT-derived storage at the same date. Both experiments use identical future inflow ensembles and identical parameters for the reservoir model. Their difference quantifies the effect of replacing contemporaneous in-situ storage with the SWOT-based initial state, including any error introduced while propagating the latest SWOT observation to T0.

This approach preserves a regular operational forecast schedule rather than restricting forecasts to SWOT overpass dates and is consistent with the monthly initialization schedule of GloFAS. Initialization effects are expected to be strongest near T0, whereas accumulated inflow uncertainty becomes increasingly important with lead time; however, reservoir storage memory may allow initial-state effects to persist.

4.4.4 Hanasaki Reservoir Operation Model

We simulated reservoir storage and release with the Hanasaki reservoir operation model (Hanasaki et al., 2006). For each ensemble member, the model applies the reservoir water balance and operating rules to propagate the initialized storage under the forecast inflow trajectory. Reservoir-specific storage capacity, elevation-storage information, mean annual inflow, and demand or release characteristics are supplied where available, while the general model structure was unchanged across BUL, DNP, and Tendaho.

For each WUS forecast cycle, we ran the model for four combinations: observed-storage initialization with ESP inflow, SWOT-based initialization with ESP inflow, observed-storage initialization with GloFAS inflow, and SWOT-based initialization with GloFAS inflow. We simulated each ensemble member independently and summarized the resulting storage and release trajectories across members. Forecasts are issued monthly and continue through the available forecast horizon, up to the end of the water year where possible.

This factorial design separates the influence of the initial storage source from that of the inflow forecast source. At Tendaho, we applied the same model sequence using SWOT initialization and bias-corrected GloFAS inflow, with results interpreted in light of the more limited observational record.

4.4.5 Forecast Evaluation and Agricultural Demand-Satisfaction Assessment

We evaluated forecast performance for inflow, storage, release, and agricultural-demand satisfaction. We compared inflow forecasts with observed inflow by issue date and lead time. We evaluated storage forecasts relative to observed storage. Paired simulations quantify the effect of replacing observed-storage initialization with SWOT-based initialization. We compared release forecasts with observed release where available. We assessed water-management relevance by directly comparing forecast-implied agricultural water supply with estimated monthly agricultural demand. Forecasted supply below the demand estimate indicates a potential unmetdemand condition, whereas supply above demand indicates that the estimated agricultural requirement could be met under the forecasted release scenario.

4.5 Results

4.5.1 SWOT-Based Reservoir Storage Initialization

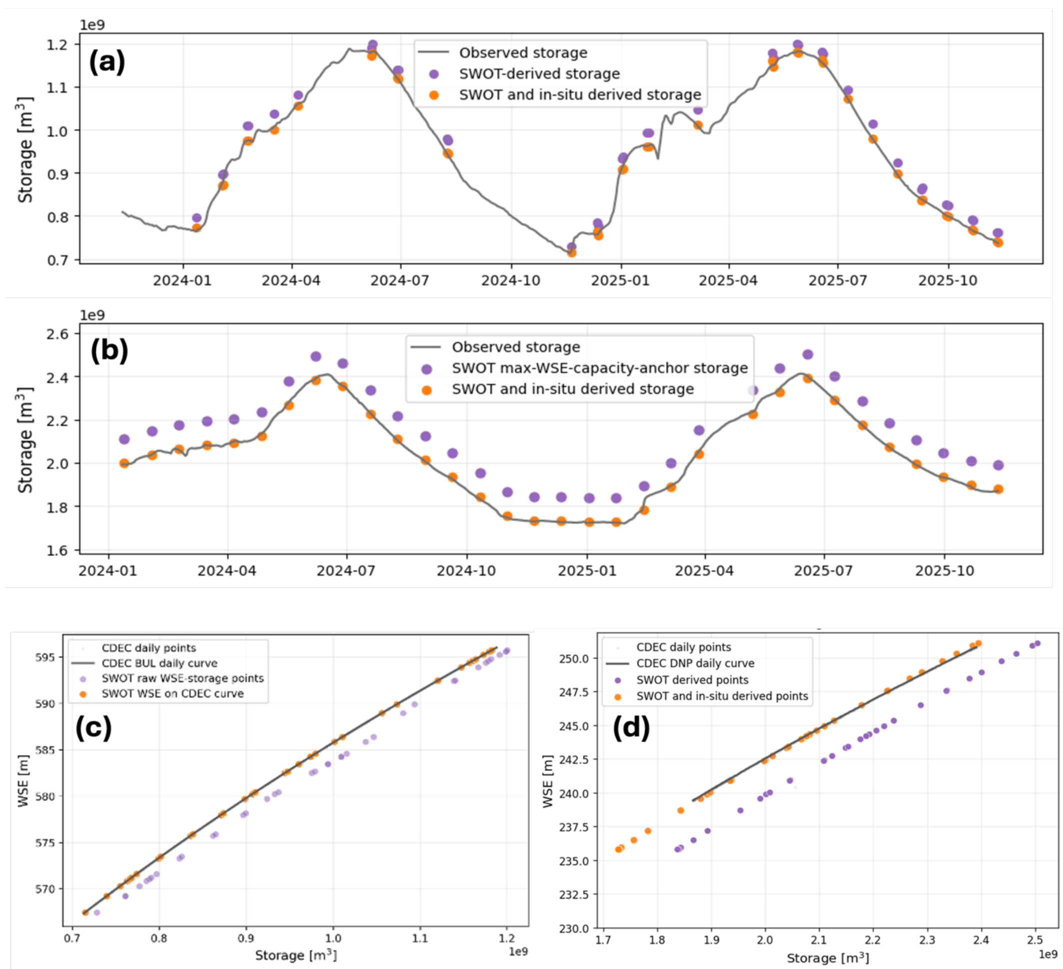


Figure 4.3. Comparison of SWOT-derived and in-situ reservoir storage and elevation–storage relationships for New Bullards Bar and Don Pedro reservoirs. Panels (a) and (b) show observed storage, storage derived directly from SWOT observations, and storage estimated by combining SWOT and in-situ information for New Bullards Bar and Don Pedro,

respectively. Panels (c) and (d) compare the water-surface elevation–storage relationships for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP), respectively, using agency records, SWOT-derived observations, and combined SWOT–in-situ reference points.

Figure 4.3 compares the storage time series derived from SWOT with the in-situ storage records for BUL and DNP. For BUL, the combined SWOT–in-situ storage estimates closely follow the observed seasonal filling and drawdown, including the two spring refill periods and subsequent summer declines. The direct SWOT-derived estimates show somewhat greater scatter, but they generally preserve the timing and magnitude of the observed seasonal cycle. The mean absolute difference between the combined estimates and the observed record is 25.1 million m3, compared with 3.5 million m3 for the direct SWOT estimates.

For DNP, the maximum-WSE capacity anchor (defined by pairing the maximum quality-controlled SWOT-observed WSE during the science phase with the reservoir's reported design maximum storage capacity) produces a persistent positive storage offset relative to the observed record, whereas the combined SWOT–in-situ estimates more closely reproduce the observed seasonal variation. The corresponding elevation–storage comparisons in Figures 4.3c and 4.3d show that the combined reference points align more closely with the agency curves than the unadjusted SWOT-derived points.

Across the two reservoirs, the SWOT-only storage estimates capture 98.5% of the observed seasonal storage range at BUL and 98.2% at DNP. Agreement is strongest under filling/refill conditions and weakest under drawdown conditions. The larger persistent offset at DNP likely reflects the fact that the reservoir reached only 95.3% of its reported maximum storage capacity during the SWOT science phase, based on the in-situ record. Consequently, the maximum SWOT-observed WSE does not represent the true full-pool elevation and provides an imperfect upper anchor for the elevation–storage relationship. For reservoirs that do not approach full capacity during the satellite record, additional reservoir information, such as in-situ WSE and storage observations, an independently established elevation–storage curve, or reliable design information linking full-pool elevation to maximum storage capacity, may be necessary to derive accurate absolute storage estimates. In data-limited regions, SWOT-derived storage may be more reliable for reservoirs with high inflow-to-capacity ratios, such as BUL, because their higher turnover increases the likelihood that the satellite record captures near-full conditions and therefore provides a more representative upper elevation–storage anchor. More generally, storage estimates may also improve as the SWOT observational record lengthens, because a longer record increases the probability of observing reservoir states closer to full capacity.

Overall, the storage comparison establishes that SWOT observations contain sufficient information to track the seasonal reservoir state, but it also shows that reservoir-specific elevation–storage anchoring remains an important source of systematic error. The influence of these biases on forecast initialization and subsequent storage predictions is examined in the following section (Section 4.5.3).

4.5.2 Seasonal Reservoir Inflow Forecasts

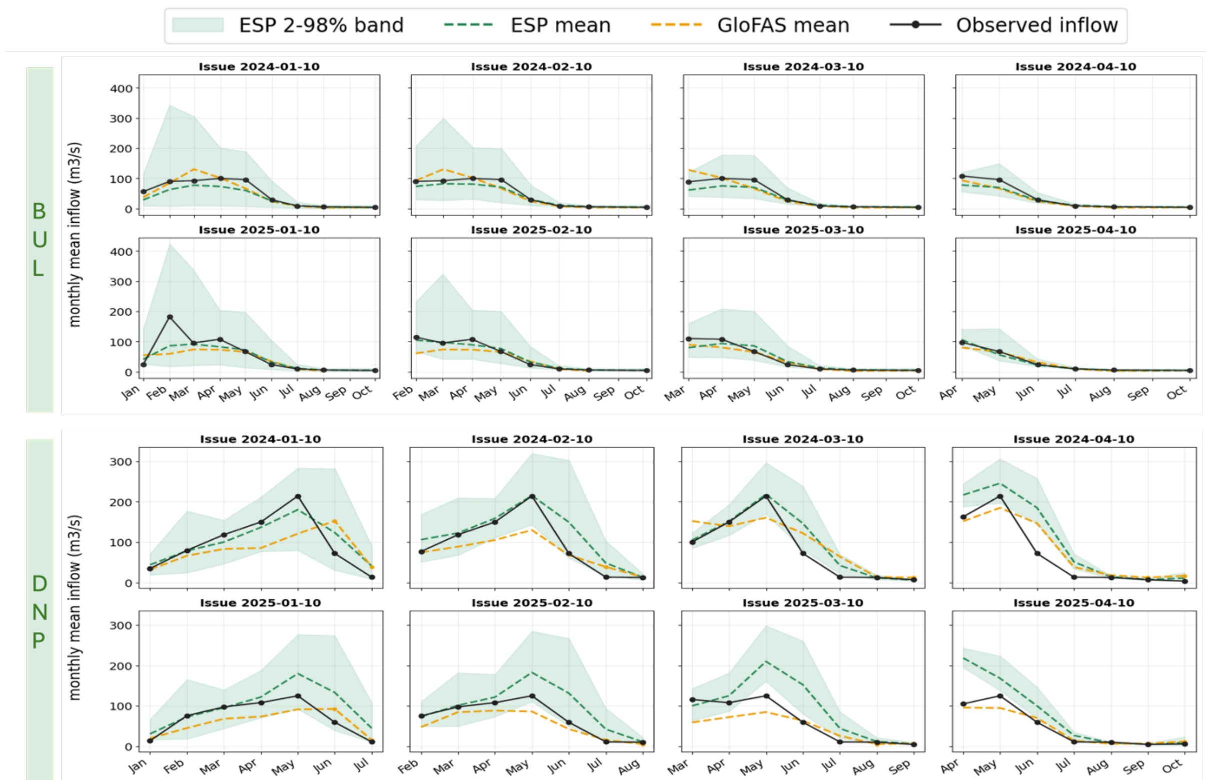


Figure 4.4. Seasonal reservoir inflow forecasts from ESP and GloFAS for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP) across monthly forecast issue dates.

Figure 4.4 compares ESP and GloFAS inflow forecasts with observed reservoir inflow across monthly issue dates. Both forecast systems reproduce the broad seasonal transition from winter–spring high flows to low summer inflows at BUL and DNP. Forecast spread is largest during the high-flow and snowmelt periods and narrows as observed inflow approaches zero later in the water year.

For BUL, ESP generally follows the observed seasonal trajectory more closely during the early- and mid-season forecast cycles, whereas GloFAS shows larger departures during several high-flow months. For DNP, both products capture the timing of the spring inflow peak, but the magnitude of the peak and the subsequent decline vary among issue dates. Across all cycles, the mean normalized inflow error is 0.22 for ESP and 0.26 for GloFAS at BUL, and 0.23 and 0.38, respectively, at DNP.

A cycle-by-cycle comparison shows that ESP has the smaller inflow error in 4 of 8 BUL forecast cycles and 8 of 8 DNP cycles. The advantage is concentrated in the winter–spring high-flow season, when the magnitude and timing of snowmelt and storm-driven inflows exert the greatest control on the seasonal water balance. During the low-flow summer period, the absolute difference between ESP and GloFAS decreases because observed inflow approaches zero; consequently, smaller summer errors should not necessarily be interpreted as greater forecast skill. A detailed comparison of the May–August forecast cycles is provided in Figures B3 and B4 (Section S2) for BUL and DNP, respectively.

The ensemble range expands most rapidly over the first two lead months at BUL and over the first four to five lead months at DNP, then contracts as members converge toward low summer inflow. Across all lead times, median absolute inflow error decreases from 17.8 m3/s at leads of 0–30 days to 4.0 m3/s at leads longer than 90 days for ESP, compared with 25.6 and 3.2 m3/s for GloFAS. This decline reflects the transition into low-flow summer conditions rather than an improvement in long-lead forecast skill.

Taken together, the inflow results indicate that the locally configured ESP system provides the more accurate forcing during hydrologically active periods, while GloFAS retains the broad seasonal signal and therefore provides a viable global alternative where a local ESP implementation is unavailable.

4.5.3 Reservoir Storage Forecasts

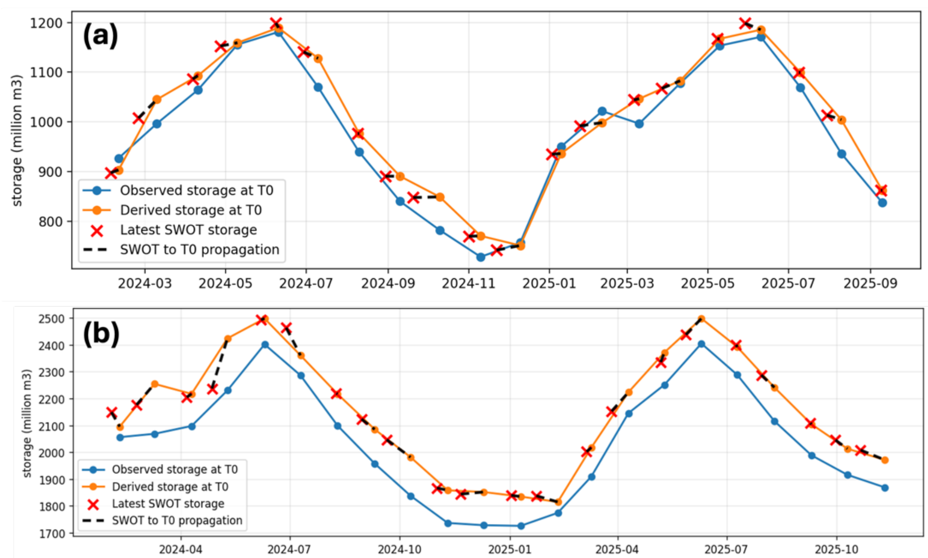


Figure 4.5. Comparison of observed and SWOT-based reservoir storage at monthly forecast issue dates for New Bullards Bar Reservoir (a; BUL) and Don Pedro Reservoir (b; DNP). Blue circles indicate observed (in situ) storage at the forecast initialization date (T0), red crosses indicate the latest available SWOT-derived storage observations before T0, black dashed lines show their propagation to T0, and orange circles indicate the resulting propagated SWOT-based storage estimates at T0. The propagation procedure is described in Section 4.4.3.

Figure 4.5 evaluates the storage state used at each monthly forecast issue date. At BUL, the propagated SWOT-based storage generally follows the observed storage at T0 and reproduces the seasonal refill and drawdown pattern, although small positive and negative deviations occur among issue dates. The mean absolute initialization difference is 62.4 million m3.

At DNP, the propagated SWOT-based storage is consistently higher than the observed storage over most of the study period, reflecting the positive offset in the SWOT-derived storage relation shown in Figure 4.3. Despite this bias, the propagated initialization retains the observed timing of seasonal storage changes. The corresponding mean initialization bias is 110.8 million m3.

The contrast between BUL and DNP illustrates the reservoir-specific persistence of initialization error. At BUL, the sign of the initialization difference changes among issue dates, and the mean signed difference remains modest (+30.5 million m3, or +3.1% of mean observed T0 storage). At DNP, the difference is consistently positive, with a mean bias of +110.8 million m3 (5.4%). The initialization error remains above 25 million m^3 for 22 consecutive issue dates at DNP, consistent with the persistent offset in the elevation-storage relationship.

Despite these differences, the propagated SWOT state captures the phase of the seasonal storage cycle at every issue date. Thus, the principal limitation of the satellite-based initialization is the magnitude of storage rather than the timing of reservoir filling and drawdown.

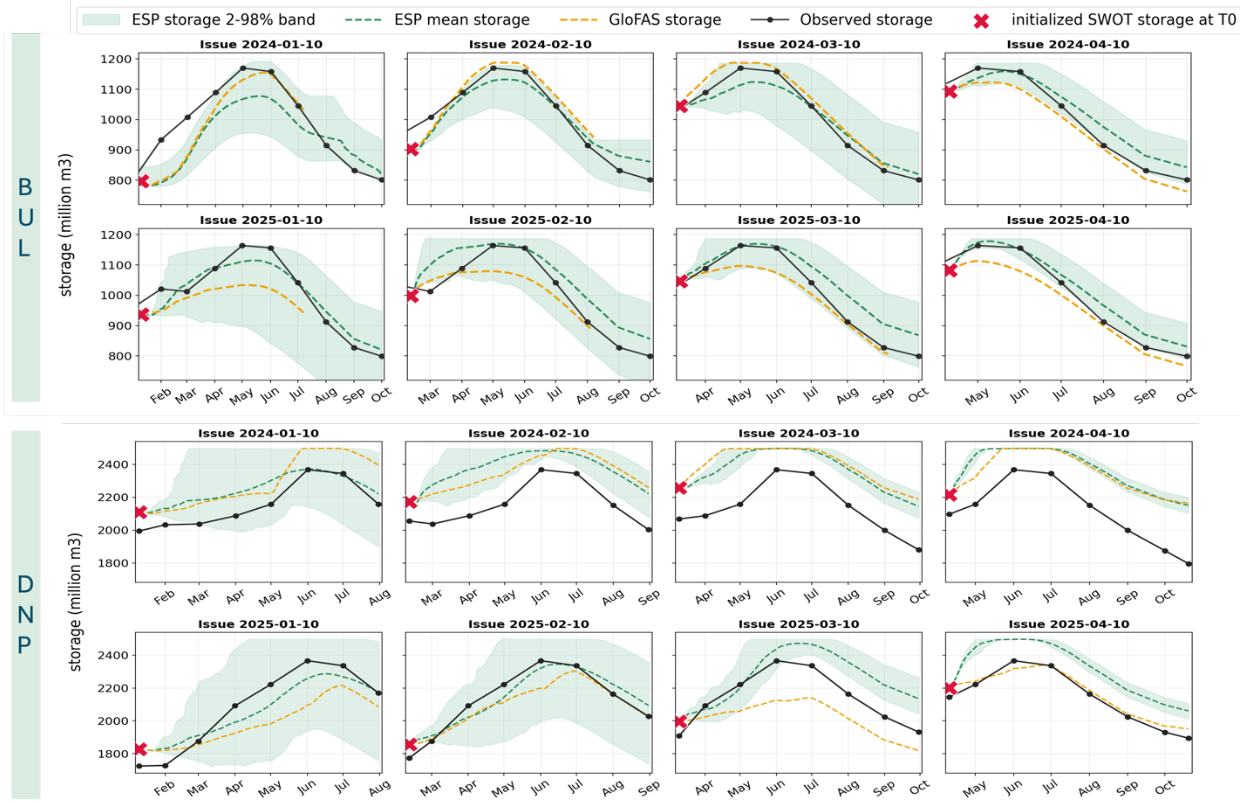


Figure 4.6. Comparison of ESP- and GloFAS-driven reservoir storage forecasts initialized with SWOT-based storage in comparison with observed (in situ) storage for New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP) in 2024 and 2025. Black solid lines with circles represent observed storage, orange dashed lines represent forecasts driven by ESP inflows, green dashed lines represent forecasts driven by GloFAS seasonal inflows, green shading indicates the GloFAS forecast uncertainty range, and red crosses identify the SWOT-based initial storage at each forecast issue date. Each panel corresponds to an individual monthly forecast issue date.

The storage forecasts in Figure 4.6 show that both inflow products reproduce the principal seasonal pattern of spring filling followed by summer drawdown. For BUL, ESP-driven forecasts generally remain closer to the observed trajectory during the refill period, while the

GloFAS-driven forecasts show larger departures for several forecast cycles. Differences between the two forecast products become smaller during the late-summer drawdown, when inflow is low, and reservoir storage declines more steadily.

For DNP, ESP-driven forecasts capture the seasonal rise and peak more consistently across the displayed years, whereas GloFAS-driven forecasts show greater underestimation during refill and larger spread among forecast cycles. However, both products reproduce the timing of the summer storage decline. Across all issue dates, the mean storage MAE is 42 million m^3 for ESP and 80 million m^3 for GloFAS at BUL, and 142 and 143 million m^3, respectively, at DNP.

The effect of inflow forecast source becomes more pronounced as forecast lead increases. At short leads of 0–30 days, storage trajectories initialized from the same state remain relatively similar, whereas after 90 days the separation between ESP- and GloFAS-driven forecasts increases to an average of 99 million m^3 at BUL and 105 million m^3 at DNP. This pattern indicates a transition from initial-condition control near the issue date to accumulated inflow control later in the forecast horizon.

Interannual differences are also evident, with larger errors in 2024 than in 2025 at DNP, particularly during the 2024 refill season, when forecast inflow errors accumulate rapidly. Conversely, during the late-summer drawdown, the reservoir trajectories become smoother and the spread among inflow-driven forecasts decreases.

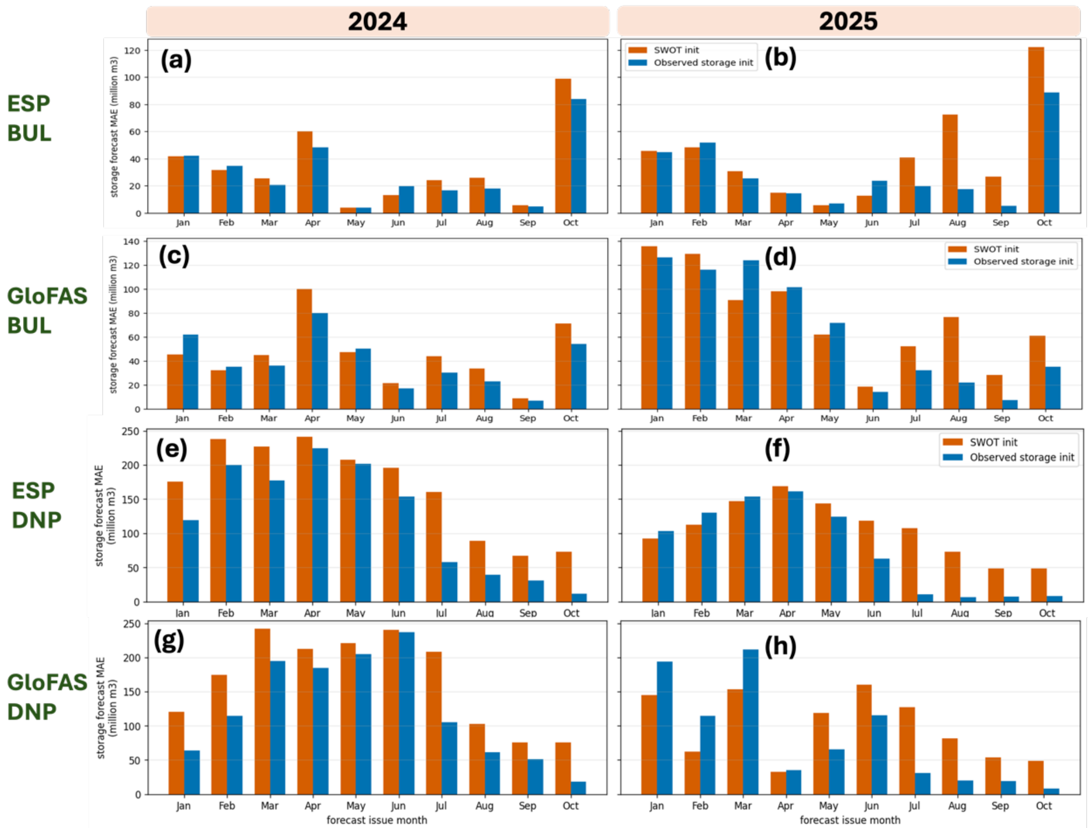


Figure 4.7. Monthly reservoir storage forecast errors under SWOT-based and observed-storage initialization for BUL and DNP using ESP and GloFAS inflow forecasts in 2024 and 2025.

Figure 4.7 compares monthly storage forecast errors obtained using SWOT-based and observed-storage initialization. At BUL, the two initialization approaches generally produce comparable storage MAEs, although observed-storage initialization yields lower errors for most issue months. Differences are relatively small for many forecast cycles, indicating that the propagated SWOT storage at T0 is often close to the observed initial state. Larger discrepancies occur for several late-season issue dates, most notably August 2025 under the GloFAS forecasts, when the SWOT-initialized error substantially exceeds that obtained using observed storage.

At DNP, the influence of the initialization source is larger and more persistent. SWOT-based initialization generally produces higher storage MAEs than observed-storage initialization across both ESP- and GloFAS-driven forecasts, consistent with the positive bias in SWOT-derived storage identified in Figures 4.3 and 4.5. The difference is particularly evident for several issue months in 2024 and early 2025. Nevertheless, the magnitude of the forecast error also varies substantially between ESP and GloFAS, showing that inflow forecast uncertainty remains an important contributor to total storage error. Averaged across all forecast cycles, replacing observed storage with SWOT-based initialization changes storage MAE by approximately 19.5% for BUL and 34.3% for DNP.

Across the complete set of monthly issue dates, SWOT-based initialization produces lower storage MAE than observed-storage initialization in 20 of the 80 experiments and higher MAE in 60 experiments. Thus, although observed-storage initialization generally provides the more accurate benchmark, SWOT-based initialization remains comparable for a meaningful subset of forecast cycles. Its performance is strongest when the satellite-derived initial state is close to the observed storage, whereas persistent bias in the SWOT estimate, as at DNP, degrades forecast accuracy across successive issue dates.

The relative contributions of initialization and inflow uncertainty also vary with forecast lead time. Near T0, the difference between paired initialization experiments is approximately 40 million m^3, whereas the difference between ESP- and GloFAS-driven forecasts is approximately 5 million m^3. At lead times longer than 60 days, the difference associated with the inflow product increases to approximately 30 million m^3 and exceeds the initialization contribution in about 70% of the experiments. These results indicate that initial-storage uncertainty is most influential early in the forecast, while inflow uncertainty increasingly controls storage accuracy at longer seasonal lead times.

Overall, Figures 4.5–7 demonstrate that SWOT-derived storage can provide a practical alternative to contemporaneous in situ storage for forecast initialization, particularly when the satellite-derived storage has little systematic bias. However, the contrast between BUL and DNP shows that persistent bias in the SWOT-derived initial state can propagate through multiple forecast cycles. At longer lead times and during periods of substantial inflow, the accuracy of the seasonal inflow forecast becomes the dominant control on storage prediction performance.

4.5.4 Forecasted Agricultural Water Supply Relative to Estimated Demand

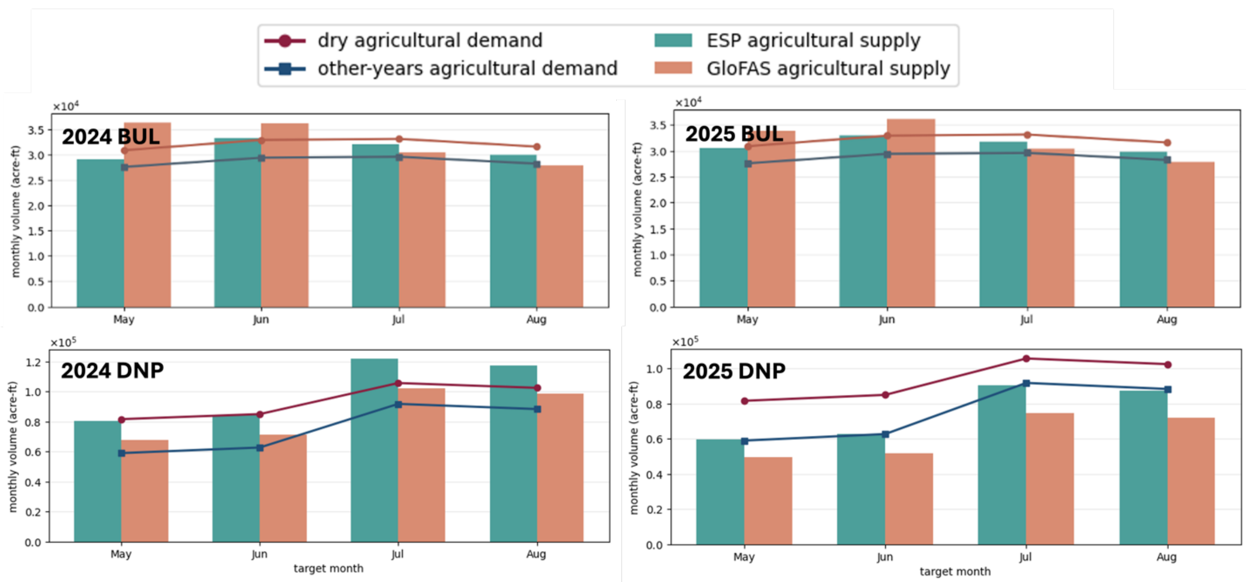


Figure 4.8. May–August agricultural water-supply derived from monthly initialized reservoir simulations forced by ESP and GloFAS inflow forecasts, compared with irrigation-demand estimates for other (all) years and dry years. Dry years are defined as the five driest water years during the agriculture record period (water year 1980–2020 for BUL; year 2012-2023 for DNP).

Figure 4.8 compares forecast-implied agricultural water supply with the estimated monthly demand benchmarks. Across the displayed May–August target months, ESP generally produces higher agricultural supply estimates than GloFAS. In panels (a) and (b), both products are near or above the other-years demand benchmark in most months, while neither product consistently reaches the higher dry-year demand benchmark.

In panels (c) and (d), the contrast between the inflow products is larger. ESP-derived supply is generally closer to, or exceeds, the corresponding demand estimates, whereas GloFAS-derived supply falls below demand in several months, especially during July and August. These comparisons indicate that conclusions about potential demand satisfaction depend on both the seasonal inflow forecast and the assumed hydrologic-year demand scenario. The average supply–demand difference is approximately +10,000 acre-ft for ESP and −6,000 acre-ft for GloFAS.

The comparison also reveals a clear interannual contrast. Forecast-implied agricultural supply is higher in 2024 than in 2025 for both inflow products, with the largest year-to-year decrease occurring in May. Under the other-years demand benchmark, forecasted supply is at or above demand in 13 of the 16 displayed month–year combinations for ESP and 7 of 16 combinations for GloFAS. Under the dry-year benchmark, the corresponding counts decrease to 8 and 7, respectively.

Differences attributable to storage initialization are smaller than those attributable to the inflow product in all 8 of the displayed target months. The mean absolute difference between supply estimates generated from SWOT- and observed-storage initialization is approximately 6,000 acre-ft, compared with an ESP–GloFAS difference of approximately 19,000 acre-ft. This indicates that agricultural supply estimates are relatively robust to the use of SWOT-based initialization, supporting its reliability for reservoir-based supply assessment.

Potential supply–demand mismatches are concentrated in July–August, when irrigation demand remains elevated but forecast-implied reservoir inflow and release begin to decline. In contrast, supply above the estimated demand during May–June should be interpreted as excess forecast-implied release availability rather than as additional realized agricultural use. The results therefore describe the capacity of forecasted releases to meet the demand benchmark, not actual farm-level water deliveries.

Overall, the management assessment is sensitive to both the selected inflow forecast and the assumed demand scenario. ESP-driven simulations more frequently meet the estimated demand benchmarks, whereas the lower GloFAS-driven supply in several months produces a more conservative indication of potential demand satisfaction.

4.5.5 Application of the Validated Workflow in a Data-Limited Region: Tendaho Reservoir

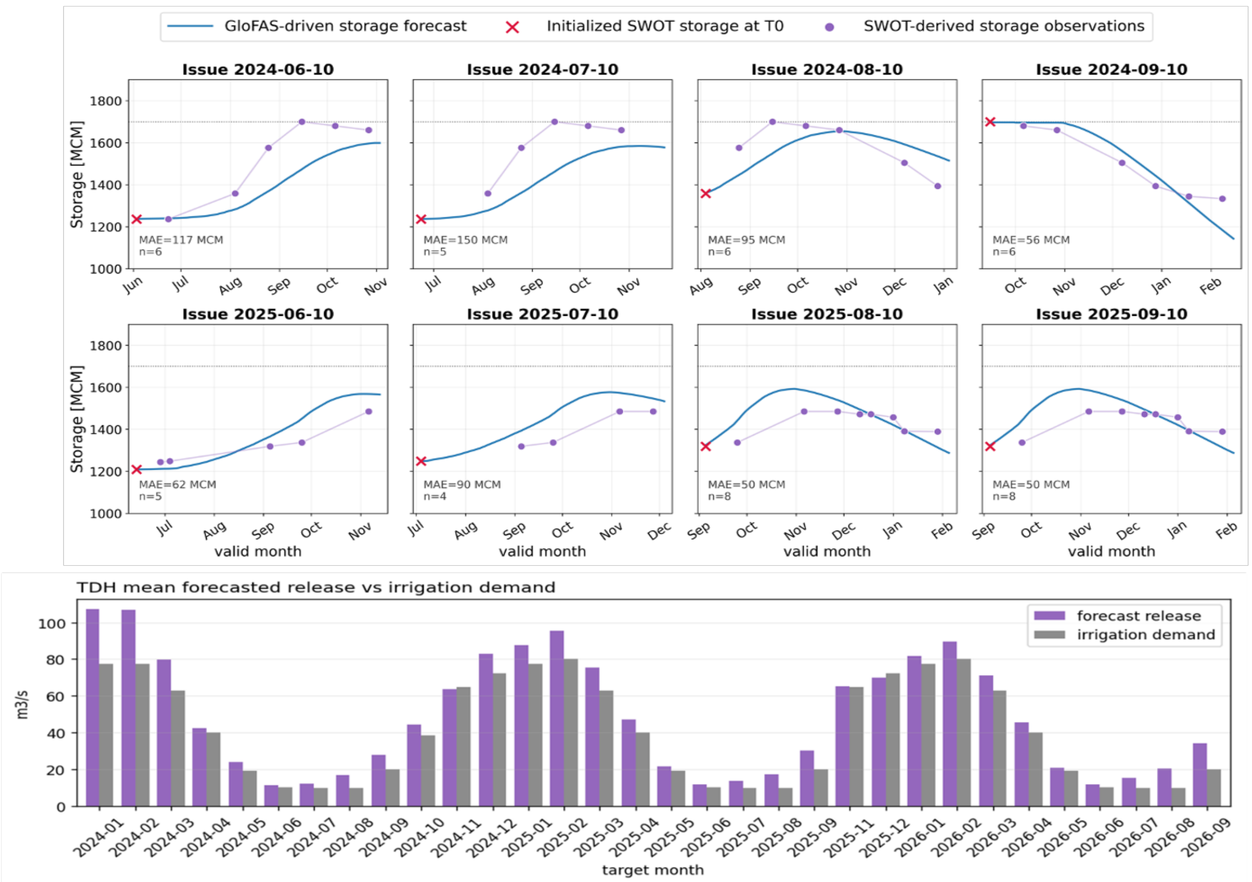


Figure 4.9. GloFAS-driven seasonal reservoir forecasts for Tendaho Reservoir. The upper panels show forecast storage trajectories initialized with SWOT-derived storage at T0 for eight issue months in 2024 and 2025, while the lower panel compares monthly mean forecast release with estimated irrigation demand.

Figure 4.9 shows the transferred workflow for Tendaho Reservoir using SWOT-derived initial storage and bias-corrected GloFAS inflow. The SWOT-observed WSE and corresponding derived storage time series used to characterize reservoir storage conditions are shown in Figure B5 in section S3. The upper panels present storage forecasts initialized at eight issue dates during 2024 and 2025. The forecast trajectories generally show seasonal storage accumulation followed by drawdown, although the timing and magnitude of the maximum storage vary among issue dates. Across the eight forecasts, simulated peak storage ranges from approximately 1,550 to 1,700 MCM, while the lowest end-of-forecast storage is approximately 200 MCM for the September 2024 initialization.

The monthly comparison indicates that mean forecast release is generally sufficient to meet the estimated irrigation demand. Across the 33 target months from January 2024 through September 2026, forecast release exceeds demand in approximately 30 months, falls below demand in two months, and is approximately equal to demand in one month. Potential shortfalls are limited primarily to November 2024 and November 2025, with the largest monthly deficit occurring in November 2025 at approximately 3 $m^3\ s^{-1}$. In contrast, the largest positive release–demand differences occur in January and February 2024 and are approximately 30 $m^3\ s^{-1}$.

The seasonal release pattern broadly follows the prescribed irrigation-demand cycle, with the largest forecast releases generally occurring during November–February and the smallest during May–July. During months when forecast release exceeds demand, the mean positive difference is approximately 9–10 $m^3\ s^{-1}$, whereas the average shortfall during below-demand months is approximately 2 $m^3\ s^{-1}$. These results suggest that the modeled reservoir can satisfy the prescribed irrigation demand during most months, although small potential deficits may occur under particular combinations of storage, inflow, and seasonal demand.

Because continuous observed release and reservoir-specific operating-rule information are unavailable, these results should be interpreted as a demonstration of potential irrigation supply conditions rather than a direct validation of operational performance. Relative to the western U.S. cases, the Tendaho application relies more heavily on globally available and approximate inputs, including SWOT-derived initial storage, bias-corrected GloFAS inflow, generalized reservoir operating rules, and literature-based irrigation demand.

Despite these limitations, the Tendaho application preserves the same physical sequence identified in the validation reservoirs: satellite observations constrain the initial storage state, seasonal inflow forecasts determine the subsequent storage evolution, and the reservoir model translates these trajectories into a management-relevant comparison with irrigation demand. The results therefore demonstrate the potential utility of the framework for diagnosing seasonal irrigation supply conditions in regions where continuous operational observations are unavailable.

4.6 Conclusions

We developed and evaluated a SWOT-initialized reservoir-operation forecasting framework that combines SWOT-based satellite-derived reservoir storage, seasonal inflow forecasts, reservoir-operation modeling, and agricultural water-demand assessment. We first evaluated the framework at two data-rich California reservoirs and then applied it to Tendaho Reservoir in Ethiopia to demonstrate its applicability in a data-limited setting.

Our results show that SWOT-derived storage can provide a practical alternative to in situ storage observations for reservoir-operation forecasting at the forecast initialization date. At both California reservoirs, the SWOT-based initial states captured the timing of seasonal filling and drawdown. However, forecast performance depended strongly on the accuracy of the reservoir-specific elevation–storage relationship. Where the SWOT-derived storage contained little systematic bias, as at BUL, forecasts initialized with SWOT storage were often comparable to those initialized with observed storage. In contrast, the persistent positive storage offset at DNP propagated into subsequent forecasts and increased storage forecast errors across multiple issue dates.

The paired initialization and inflow experiment also clarified how forecast uncertainty changes with lead time. Initial-storage uncertainty had its largest influence near the forecast issue date, whereas differences associated with ESP (which requires a local history of in situ reservoir inflows) and GloFAS (available globally) inflow forecasts became increasingly important as lead time increased. This transition indicates that improvements in satellite-based reservoir-state estimation are especially valuable for short-lead forecasts, while seasonal forecast performance increasingly depends on the accuracy of future inflow estimates, which remain dependent on in situ data.

Application to Tendaho Reservoir demonstrates the potential of combining SWOT-derived storage with bias-corrected global inflow forecasts to assess agricultural water availability where continuous reservoir observations are unavailable. The framework therefore provides a transferable approach for identifying periods of potential water shortage and evaluating demand satisfaction in data-limited regions. Nevertheless, the Tendaho results should be interpreted as a demonstration rather than a fully validated forecast application because continuous observations of storage, release, and operating rules were unavailable during the SWOT period.

Overall, the study shows that satellite observations can extend reservoir-operation forecasting beyond well-monitored systems. Continued improvements in elevation–storage relationships, longer SWOT observational records, inflow bias correction, and representation of reservoir operating rules will further enhance the reliability and operational value of our framework.

References

Ahmad, A., El-Shafie, A., Razali, S. F. M., and Mohamad, Z. S.: Reservoir optimization in water resources: a review, Water Resources Management, 28, 3391–3405, https://doi.org/10.1007/s11269-014-0700-5, 2014.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrology and Earth System Sciences, 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013.

Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and its capabilities for land hydrology, in: Remote Sensing and Water Resources, edited by: Cazenave, A., Champollion, N., Benveniste, J., and Chen, J., Springer International Publishing, Cham, 117–147, https://doi.org/10.1007/978-3-319-32449-4_6, 2016.

California Department of Water Resources: Agricultural Water Use Data 1998–2005, California Natural Resources Agency Open Data Portal, https://catalog.data.gov/dataset/agricultural-water-use-data-1998-2005-f2aa3, 2024a.

California Department of Water Resources: Agricultural Water Use Data 2006–2010, California Natural Resources Agency Open Data Portal, https://catalog.data.gov/dataset/agricultural-water-use-data-2006-2010-60e9a, 2024b.

California Department of Water Resources: Agricultural Water Use Data 2011–2015, California Natural Resources Agency Open Data Portal, https://catalog.data.gov/dataset/agricultural-water-use-data-2011-2015, 2024c.

California Department of Water Resources: Statewide Agricultural Water Use Data 2016–2020, California Natural Resources Agency Open Data Portal, https://catalog.data.gov/dataset/statewide-agricultural-water-use-data-2016-2020-313ba, 2024d.

California Department of Water Resources: WUEdata Portal – Aggregated Farm-Gate Delivery Reports, https://wuedata.water.ca.gov/ (last access: 15 August 2026), n.d.

California Department of Water Resources: California Irrigation Management Information System (CIMIS), Spatial CIMIS reference evapotranspiration data, https://cimis.water.ca.gov/SpatialData.aspx (last access: 15 August 2026), n.d.

Day, G. N.: Extended streamflow forecasting using NWSRFS, Journal of Water Resources Planning and Management, 111, 157–170, https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Delaney, C. J., Hartman, R. K., Mendoza, J., Dettinger, M., Delle Monache, L., Jasperse, J., Ralph, F. M., Talbot, C., Brown, J., Reynolds, D., and Evett, S.: Forecast informed reservoir operations using ensemble streamflow predictions for a multipurpose reservoir in Northern California, Water Resources Research, 56, e2019WR026604, https://doi.org/10.1029/2019WR026604, 2020.

Dettinger, M. D. and Anderson, M. L.: Storage in California's reservoirs and snowpack in this time of drought, San Francisco Estuary and Watershed Science, 13, https://doi.org/10.15447/sfews.2015v13iss2art1, 2015.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geoscientific Model Development, 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Fu, L.-L., Pavelsky, T., Crétaux, J.-F., Morrow, R., Farrar, J. T., Vaze, P., Sengenes, P., Vinogradova-Shiffer, N., Sylvestre-Baron, A., and Picot, N.: The Surface Water and Ocean Topography Mission: a breakthrough in radar remote sensing of the ocean and land surface water, Geophysical Research Letters, 51, e2023GL107652, https://doi.org/10.1029/2023GL107652, 2024.

Gao, H., Birkett, C., and Lettenmaier, D. P.: Global monitoring of large reservoir storage from satellite remote sensing, Water Resources Research, 48, W09504, https://doi.org/10.1029/2012WR012063, 2012.

Gochis, D. J., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., and Zhang, Y.: The

WRF-Hydro® modeling system technical description, Version 5.1.1, NCAR Technical Note, 2020.

Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, Journal of Hydrology, 327, 22–41, https://doi.org/10.1016/j.jhydrol.2005.11.011, 2006.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing, Hydrology and Earth System Sciences, 12, 1007–1025, https://doi.org/10.5194/hess-12-1007-2008, 2008.

Iglesias, M. A., Law, K. J. H., and Stuart, A. M.: Evaluation of Gaussian approximations for data assimilation in reservoir models, Computational Geosciences, 17, 851–885, https://doi.org/10.1007/s10596-013-9359-x, 2013.

Joint Research Centre and Copernicus Emergency Management Service: Seasonal forecasts of river discharge and related data by the Global Flood Awareness System, Early Warning Data Store, https://doi.org/10.24381/cds.00b6c4fb (last access: 16 July 2026), 2020.

Kim, S.-J., Bae, S.-J., Lee, S.-J., and Jang, M.-W.: Monthly agricultural reservoir storage forecasting using machine learning, Atmosphere, 13, 1887, https://doi.org/10.3390/atmos13111887, 2022.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world’s reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494–502, https://doi.org/10.1890/100125, 2011.

Ma, X., Hartman, R. K., Pan, M., and Lettenmaier, D. P.: Model Calibration and Data Assimilation for the National Water Model: A Case Study over California, Journal of Hydrometeorology, https://doi.org/10.1175/JHM-D-25-0066.1, 2026a.

Ma, X., Wang, J., Gharari, S., Mizukami, N., and Lettenmaier, D. P.: Monitoring Reservoir Storage Using SWOT Satellite Observations and a Reservoir Operation Model, Water Resources Research, 62, e2025WR041223, https://doi.org/10.1029/2025WR041223, 2026b.

Maddu, R., Pradhan, I., Ahmadisharaf, E., Singh, S. K., and Shaik, R.: Short-range reservoir inflow forecasting using hydrological and large-scale atmospheric circulation information, Journal of Hydrology, 612, 128153, https://doi.org/10.1016/j.jhydrol.2022.128153, 2022.

McGuire, M., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P.: Use of satellite data for streamflow and reservoir storage forecasts in the Snake River Basin, Journal of Water Resources Planning and Management, 132, 97–110, https://doi.org/10.1061/(ASCE)0733-9496(2006)132:2(97), 2006.

Ministry of Irrigation and Lowlands: Tendaho Dam and Irrigation Development Project, Afar National Regional State, Ethiopia, https://mills.gov.et/portfolio/tendaho-dam-and-irrigation-development-project/ (last access: 15 August 2026), n.d.

NASA/JPL PO.DAAC: SWOT Level 2 Lake Single-Pass Vector Data Product, Version D (SWOT_L2_HR_LakeSP_D), Physical Oceanography Distributed Active Archive Center, Jet Propulsion Laboratory, Pasadena, California, https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_LakeSP_D (last access: 15 August 2026), 2026.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116, D12109, https://doi.org/10.1029/2010JD015139, 2011.

Özdoğan-Sarıkoç, G., Sarıkoç, M., Celik, M., and Dadaser-Celik, F.: Reservoir volume forecasting using artificial intelligence-based models: Artificial Neural Networks, Support Vector Regression, and Long Short-Term Memory, Journal of Hydrology, 616, 128766, https://doi.org/10.1016/j.jhydrol.2022.128766, 2023.

Pan, M.: CW3E 1-km 1-hourly Meteorological Forcing on NWM Grid [data set], Center for Western Weather and Water Extremes (CW3E), https://doi.org/10.5281/zenodo.14714512, 2025.

Rahmanifard, H. and Gates, I.: A comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: best practices and future directions, Artificial Intelligence Review, 57, 213, https://doi.org/10.1007/s10462-024-10865-5, 2024.

Shonka, T. G. and Gullie, G.: Determining the optimum decision variables and rescheduling the irrigation system of Tendaho Sugar Estate, Ethiopia, Irrigation & Drainage Systems Engineering, 6, 200, https://doi.org/10.4172/2168-9768.1000200, 2017.

Spinti, R. A., Condon, L. E., and Zhang, J.: The evolution of dam induced river fragmentation in the United States, Nature Communications, 14, 3820, https://doi.org/10.1038/s41467-023-39194-x, 2023.

Su, L., Pan, M., Hartman, R. K., and Lettenmaier, D. P.: Improving National Water Model Flood Forecast Skill over Coastal Western U.S. River Basins, Journal of Hydrometeorology, https://doi.org/10.1175/JHM-D-24-0072.1, 2025.

Sushanth, K., Mishra, A., Mukhopadhyay, P., and Singh, R.: Real-time streamflow forecasting in a reservoir-regulated river basin using explainable machine learning and conceptual reservoir module, Science of the Total Environment, 861, 160680, https://doi.org/10.1016/j.scitotenv.2022.160680, 2023.

Tian, D., He, X., Srivastava, P., and Kalin, L.: A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information, Stochastic Environmental Research and Risk Assessment, 36, 2353–2375, https://doi.org/10.1007/s00477-021-02023-y, 2022.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F.: GMES Sentinel-1 mission, Remote Sensing of Environment, 120, 9–24, https://doi.org/10.1016/j.rse.2011.05.028, 2012.

Yan, M., Huang, C., Bienstman, P., Tino, P., Lin, W., and Sun, J.: Emerging opportunities and challenges for the future of reservoir computing, Nature Communications, 15, 2056, https://doi.org/10.1038/s41467-024-45187-1, 2024.

Yang, G., Guo, S., Liu, P., and Block, P.: Sensitivity of forecast value in multiobjective reservoir operation to forecast lead time and reservoir characteristics, Journal of Water Resources Planning and Management, 147, 04021027, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001384, 2021.

Yang, S., Yang, D., Chen, J., and Zhao, B.: Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model, Journal of Hydrology, 579, 124229, https://doi.org/10.1016/j.jhydrol.2019.124229, 2019.

Zanutto, D., Ficchì, A., Giuliani, M., and Castelletti, A.: Reinforcement Learning of Multi-Timescale Forecast Information for Designing Operating Policies of Multi-Purpose Reservoirs, Water Resources Research, 61, e2023WR036724, https://doi.org/10.1029/2023WR036724, 2025.

Zhang, J., Cai, X., Lei, X., Liu, P., and Wang, H.: Real-time reservoir flood control operation enhanced by data assimilation, Journal of Hydrology, 598, 126426, https://doi.org/10.1016/j.jhydrol.2021.126426, 2021.

Zhao, T., Cai, X., and Yang, D.: Effect of streamflow forecast uncertainty on real-time reservoir operation, Advances in Water Resources, 34, 495–504, https://doi.org/10.1016/j.advwatres.2011.01.004, 2011.

Chapter 5 Conclusion

This dissertation advances new methods of hydrological prediction and observations by integrating hydrologic model improvements, satellite remote sensing, and streamflow and reservoir-operation forecasting. The three core studies reported in chapters 2-4 address connected challenges across time scales. Chapter 2 evaluates and improves short-range flood forecasts produced using the next-generation Noah-MP land surface model, which is the hydrologic core of the National Water Model. Chapters 3 and 4 evaluate water management applications of observations from the Surface Water and Ocean Topography (SWOT) satellite mission, which provides observations of surface water globally, and in particular reservoir stage and surface area which can be used to estimate reservoir storage changes. Chapter 3, in particular, evaluates the accuracy and reliability of SWOT-derived reservoir water surface elevation, surface area, and storage change estimates and demonstrates how these observations can be combined with a reservoir operation model to improve the temporal continuity of reservoir monitoring between satellite overpasses. Chapter 4 extends this observation–model integration from monitoring to prediction by incorporating SWOT-derived reservoir storage information into a seasonal reservoir-operation forecasting system that also utilizes ensemble hydrologic predictions of reservoir inflows. Together, Chapters 3 and 4 examine how satellite observations, hydrologic initial conditions, reservoir-operation models, and global forecast products can be integrated to support reservoir monitoring and water-management decisions in both data-rich and data-limited regions.

In Chapter 1, I posed three objectives that guided this research:

(1) Evaluate and improve advanced flood forecast methods through model calibration, precipitation forecast bias correction, and adjustment of initial soil moisture.

(2) Assess the accuracy of new sources of hydrologic information (in particular, SWOT water-surface elevation, surface area, and reservoir storage change estimates), and determine whether a reservoir-operation model can fill temporal gaps between satellite observations.

(3) Develop a SWOT-initialized seasonal reservoir-operation forecasting framework and evaluate its transferability from well-observed California reservoirs to a data-limited reservoir in Ethiopia.

In Chapter 2, I evaluated retrospective flood forecasts for selected flood events between 2004 and 2023 in the Russian, Yuba-Feather, and Santa Ana River basins, where the California-Nevada National Weather Service River Forecast Office has archived operational flood forecasts. These three river basins represent humid, semi-humid, and semiarid conditions in California (from north to south). I first calibrated parameters of the Noah-MP hydrologic model at six forecast locations across the three river basins. Calibration substantially improved the model's ability to reproduce flood hydrographs and peak flows as compared with default parameters (used, e.g., in the National Water Model version of Noah-MP). I then generated flood reforecasts (produced retrospectively but using data that would have been available in real time) using the same quantitative precipitation estimates and forecasts that were used to produce archived California-Nevada River Forecast Center forecasts in real time. This allowed a fair comparison of archived operational flood forecasts with the reforecasts I produced.

I found that calibrated Noah-MP produced flood reforecasts with accuracy comparable to the archived operational forecasts in the Russian and Yuba-Feather basins. Forecast accuracy generally decreased with lead time, reflecting the growing influence of precipitation forecast error. Correcting precipitation forecast biases and adjusting initial soil moisture through assimilation of observed (pre-forecast) discharge information further improved Noah-MP forecasts, especially for lead times beyond approximately two days and in the semiarid Santa Ana basin, where antecedent moisture strongly controls rainfall-runoff response. After these corrections, Noah-MP was slightly more accurate than the archived forecasts in all three river basins for several measures of flood magnitude, peak timing, and hydrograph shape. My results show that Noah-MP can provide operationally competitive flood forecasts when model parameters, forcing biases, and initial states are treated explicitly.

In Chapter 3, I evaluated SWOT observations for 12 reservoirs across California, the Upper Colorado River Basin, and the Columbia River Basin. SWOT provides simultaneous water-surface elevation and surface-area observations at approximately 12.5-day observation intervals, overcoming important hydrologic sampling limitations of earlier satellite altimeters. For approximately 19 months of observations (from August 2023 to Feb 2025), the median absolute error in SWOT water-surface elevation was less than 20 cm over the 12 reservoirs and errors for the larger California reservoirs were generally only several centimeters. When SWOT water-surface elevations were well aligned with in situ observations, the relative errors in SWOT-derived storage estimates were typically less than 5% compared with in situ storage estimates.

I found that storage observation accuracy was controlled more strongly by water-surface elevation error than by surface-area errors, and performance generally declined for smaller reservoirs or observations affected by incomplete coverage and data-quality limitations. Because valid SWOT observations are irregular in time, I assimilated available storage estimates into the Hanasaki reservoir-operation model (Hanasaki et al., 2006) to reconstruct storage between satellite overpasses. The model reproduced major filling and drawdown trajectories and remained useful even when fewer than one-quarter of valid observations were retained. Depending on the reservoir and observation scenario, model-based storage errors were generally about 5–20% with respect to in situ storage observations. Therefore, my work implies that SWOT can provide more than isolated reservoir snapshots: when combined with a reservoir model and inflow information, it can support temporally continuous estimates of reservoir storage variations.

In Chapter 4, I extended the monitoring framework from Chapter 3 to seasonal reservoir-operation forecasting. The workflow converts quality-controlled SWOT observations of water surface elevation and surface area to storage, propagates the most recent satellite-derived reservoir storage states to regular forecast dates, combines that initial state with seasonal inflow forecasts, and simulates future storage and release using the Hanasaki reservoir model (Hanasaki et al., 2006). I evaluated the framework at New Bullards Bar and Don Pedro Reservoirs in California, where observed storage, inflow, and release allowed direct comparison between SWOT-based and observed-storage initialization. I then transferred the framework to Tendaho Reservoir in Ethiopia, where continuous operational observations are limited.

The California experiments showed that SWOT-derived storage can provide useful initial conditions for seasonal forecasts. Initialization errors had their greatest effects within the first 30 days, whereas inflow uncertainty became increasingly important beyond about 60 days and was especially pronounced at lead times longer than 90 days. Ensemble Streamflow Prediction (ESP;

Day, 1985), which benefited from local historical observations and a calibrated hydrologic model, generally produced more accurate forecasts than the globally available GloFAS seasonal product. Nevertheless, GloFAS reproduced broad observed seasonal patterns of storage, release, and agricultural demand satisfaction and therefore provided meaningful information without requiring a locally calibrated streamflow forecast system, which is important at locations where in situ observations of streamflow are sparse. At Tendaho, the combination of SWOT-derived storage, GloFAS inflows, and available irrigation-demand information indicated that nominal agricultural demands were generally satisfied using my reservoir operation framework during the two-year SWOT period (WY 2024 and WY 2025), although limited observations prevented comprehensive validation.

Taken together, the three chapters establish a progression that evaluates hydrologic prediction improvements to remotely sensed state (reservoir storage) estimation and finally to a unified forecast-informed reservoir management. A common finding is that forecast performance depends on both initial conditions and future forcings. For short-range flood prediction, soil moisture initialization and precipitation forecasts strongly affect simulated response. For seasonal reservoir prediction, initial storage matters most near the forecast date, while accumulated inflow error increasingly controls longer-lead outcomes. SWOT helps address the initial-condition problem by providing globally consistent reservoir observations, while models provide temporal continuity and predictive capability that satellite observations alone cannot supply.

Several limitations of my work suggest directions for future work. Noah-MP flood forecasts should be evaluated across more basins and flood types using data-assimilation approaches that can be implemented operationally without future peak information. SWOT storage estimates would benefit from improved quality control, more robust elevation–storage relationships, explicit treatment of datum and sedimentation uncertainty, and a longer observation record. The reservoir forecasting framework could incorporate probabilistic storage initialization, ensemble inflow uncertainty, evaporation, operational constraints, and reservoir-specific release rules. Future studies could also explore artificial intelligence methods to improve both hydrological prediction and reservoir decision-making. For example, time-series models such as long short-term memory networks (LSTM) and Transformers could be used to learn nonlinear relationships among meteorological forcings, hydrologic states, reservoir storage, and release, while reinforcement learning could be applied to develop adaptive reservoir operating policies under uncertain inflow and water-demand conditions. Hybrid approaches that combine physically based models with machine learning may be especially promising because they could retain physical consistency while correcting systematic model errors and improving forecast efficiency. Applications in data-limited regions also require additional independent observations of inflow, release, irrigated area, and water demand.

Overall, this dissertation demonstrates that hydrological forecasts and reservoir-management information can be improved by combining physically based models with calibration, data assimilation, satellite observations, and global forecast products. The results support evolving streamflow prediction models such as Noah-MP for short-range flood forecasting, establish the accuracy and limitations of SWOT-based reservoir storage monitoring, and show that SWOT-derived storage can initialize seasonal reservoir-operation forecasts where conventional observations are sparse. The methods developed here provide a transferable foundation for linking Earth observations with operational water prediction and improving flood preparedness, reservoir monitoring, and water-supply assessment across diverse hydroclimatic and data environments.

Appendix A. Supplement to "Model calibration and data assimilation for the National Water Model: A case study over California"

This supplement includes Table A1, which lists the specific POT3 events for the six selected FIRO forecast points, and Figure A1, which shows the weighted average flood hydrographs from observations and Noah-MP simulations.

Table A1 POT3 events for the six selected FIRO points.

GUEC1

4-Feb-2004	18-Feb-2004	26-Feb-2004	28-Dec-2004	22-Mar-2005
19-May-2005	23-Dec-2005	1-Jan-2006	6-Mar-2006	17-Mar-2006
1-Apr-2006	13-Apr-2006	11-Feb-2007	28-Feb-2007	4-Jan-2008
26-Jan-2008	4-Feb-2008	24-Feb-2009	3-Mar-2009	4-Mar-2009
21-Jan-2010	26-Jan-2010	13-Apr-2010	30-Dec-2010	18-Feb-2011
21-Mar-2011	25-Mar-2011	20-Jan-2012	17-Mar-2012	28-Mar-2012
3-Dec-2012	24-Dec-2012	10-Feb-2014	12-Dec-2014	20-Dec-2014
7-Feb-2015	23-Dec-2015	18-Jan-2016	12-Mar-2016	16-Dec-2016
11-Jan-2017	23-Jan-2017	10-Feb-2017	21-Feb-2017	23-Mar-2018
8-Apr-2018	18-Jan-2019	15-Feb-2019	27-Feb-2019	25-Oct-2021
16-Dec-2021				

HEAC1

25-Feb-2004	22-Mar-2005	1-Mar-2006	6-Mar-2006	12-Apr-2006
10-Feb-2007	27-Jan-2008	25-Feb-2008	17-Feb-2009	24-Feb-2009
3-Mar-2009	20-Jan-2010	12-Apr-2010	20-Mar-2011	7-Mar-2016

11-Mar-2016	14-Mar-2016	11-Jan-2017	10-Feb-2017	21-Feb-2017
14-Feb-2019	27-Feb-2019	3-Mar-2019		

HOPC1

2-Jan-2004	17-Feb-2004	25-Feb-2004	28-Feb-2005	28-Mar-2005
29-Mar-2005	4-Jan-2006	1-Feb-2006	6-Mar-2006	25-Feb-2007
24-Feb-2008	17-Feb-2009	3-Mar-2009	3-Mar-2009	3-Mar-2010
3-Mar-2010	31-Dec-2010	3-Mar-2011	24-Mar-2011	9-Feb-2014
17-Jan-2016	18-Jan-2016	14-Mar-2016	22-Jan-2017	12-Feb-2017
14-Feb-2019	27-Feb-2019	2-Mar-2019		

DCSC1

24-Feb-2008	2-Mar-2009	18-Dec-2010	19-Dec-2010	19-Dec-2010
25-Feb-2011	15-Mar-2011	24-Mar-2011	16-Mar-2012	30-Nov-2012
2-Dec-2012	12-Dec-2014	21-Dec-2015	14-Mar-2016	10-Dec-2016
15-Dec-2016	10-Jan-2017	11-Jan-2017	7-Feb-2017	13-Mar-2018
22-Mar-2018	7-Apr-2018	7-Jan-2019	17-Jan-2019	26-Feb-2019
5-Apr-2020	24-Oct-2021	14-Dec-2021	23-Dec-2021	

OURC1

19-May-2005	21-May-2005	28-Dec-2005	1-Jan-2006	28-Feb-2006
3-Apr-2006	11-Feb-2007	2-Mar-2009	5-May-2009	19-Dec-2010
16-Mar-2011	17-Mar-2012	30-Nov-2012	2-Dec-2012	30-Jan-2016
10-Dec-2016	16-Dec-2016	8-Jan-2017	11-Jan-2017	9-Feb-2017
22-Mar-2018	7-Apr-2018	17-Jan-2019	14-Feb-2019	6-Mar-2019
25-Oct-2021				

YTLC1

15-Apr-2006	5-Apr-2006	28-Feb-2006	5-Jan-2008	27-Jan-2008
21-Jan-2010	20-Jan-2010	6-Feb-2010	28-Feb-2014	16-Dec-2016
24-Dec-2016	12-Jan-2017	17-Feb-2017	22-Jan-2017	2-Feb-2019

17-Jan-2019 14-Feb-2019

Main text Figure 3 presents a scatter plot comparing simulated and observed flood streamflow across six forecasting points in the FIRO watersheds. Additionally, the weighted-average flood hydrograph (Figure A1) was computed using flood peaks as weights to provide a representative flood shape. The simulated results from NWM (Figure A1) were forced by QPE, so they largely outperform the forecast results from NWM forced by QPF, as shown in main text Figure 7.

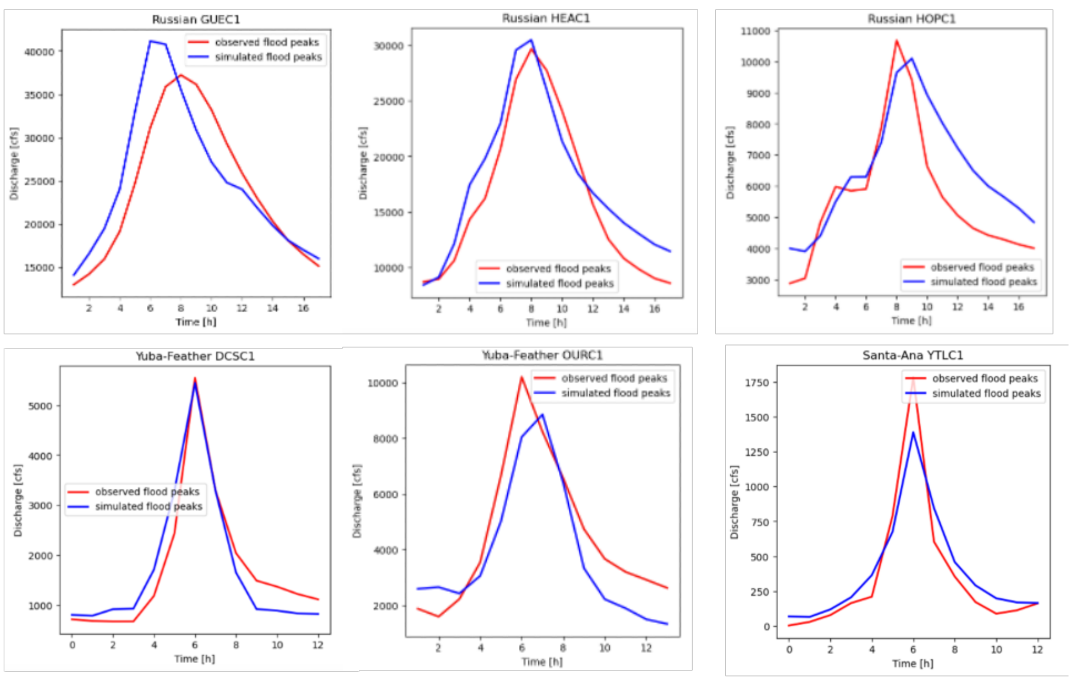


Figure A1. Weighted average flood hydrographs for observations (red) and Noah-MP simulations (blue)

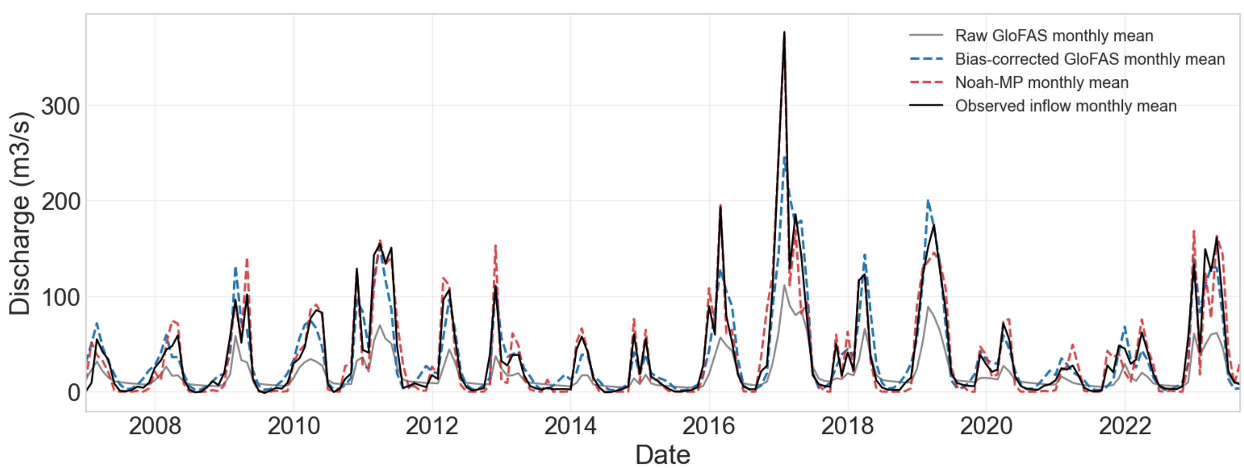
We also weighted average the different flood hydrographs using the flood peak as the weight for each flood (Figure A1). Santa Ana YTLC1 simulation (Panel 6) presents a systematic wide and short hydrograph shape, we scale the precipitation forcings use the method in 2.3.3

without changing the total amount of precipitation. The precipitation forcings for other sites' simulation keeps the original. Figure A1 illustrates the alignment between the observed and simulated flood events. Overall, there is good alignment between the observed and simulated flood curves, though slight differences are evident. In terms of peak timing, the simulated peaks generally match observed peaks closely, with minor timing shifts at some sites, such as Russian HOPC1 and Yuba-Feather OURC1, where the simulated peaks appear slightly earlier or later than observed peaks. Regarding peak magnitude, relative errors vary across sites. Before calibration, the absolute of relative errors range from 30% - 150%; while after calibration, the relative errors are 7.5%, 3.4%, 11.1%, 0.3%, 20.2%, 22.9%, respectively for GUEC1, HEAC1, HOPC1, DCSC1, OURC1, and YTLC1. The greatly improved relative errors indicate model calibration is a valid approach to enhance streamflow simulation; and such parameters, can be leveraged for flood forecasting using QPF.

Appendix B. Supplement to "SWOT-initialized reservoir operation forecasting: Validation in data-rich and application to data-limited regions"

Section S1. Evaluation of Historical Inflow Simulations and GloFAS Bias Correction

Figures S1 and S2 compare historical monthly inflow simulations from the bias-corrected GloFAS product (using the procedure described in Section 4.2 of the main text) and the Noah-MP-based modeling framework with in situ observations at New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP), respectively.. Both simulations capture the broad seasonal and interannual variability in monthly inflows, although their performance varies by reservoir and model. At BUL (Figure S1), the bias-corrected GloFAS and Noah-MP-based historical inflow simulations achieved (monthly) KGE values of 0.85 and 0.95, respectively, with corresponding RMSE values of 21.0 and 14.3 $m^3 s^{-1}$. At DNP (Figure B2), the corresponding KGE values were 0.82 and 0.85, respectively, with corresponding RMSE values of 35.3 and 40.7 $m^3 s^{-1}$. Comparison of the raw and bias-corrected GloFAS simulations in Figures S1 and S2 shows that bias correction improved the representation of seasonal peak inflows. Nevertheless, both the bias-corrected GloFAS and Noah-MP-based historical inflow simulations showed larger errors during high-flow months, as indicated by the greater scatter at higher observed inflows in Figures S1 and S2.



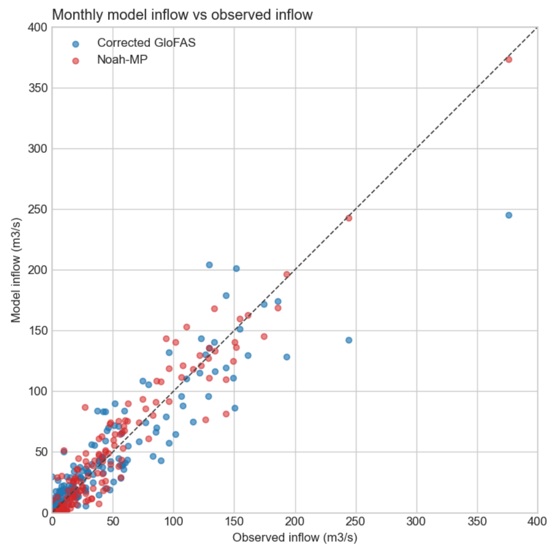
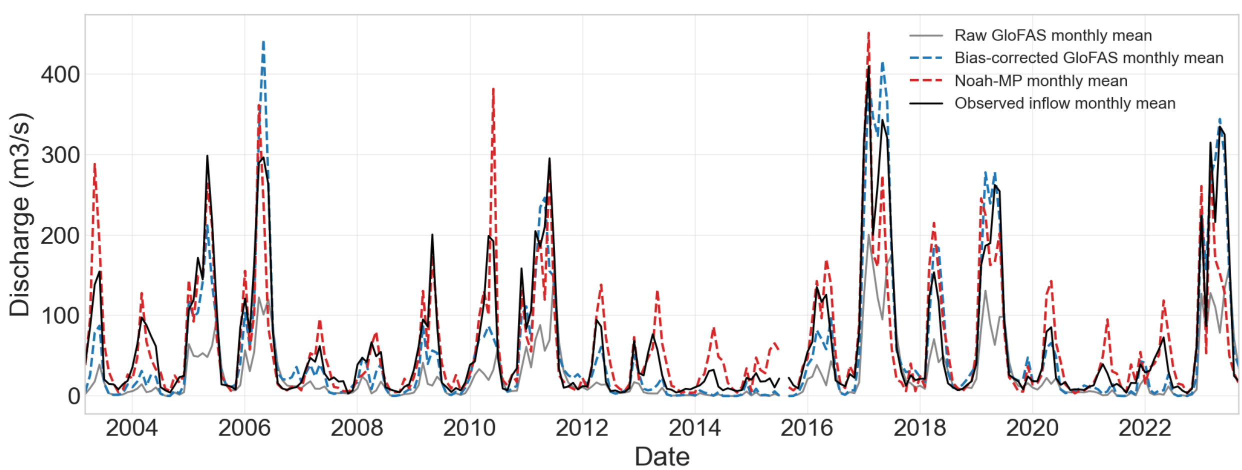


Figure S1. Evaluation of bias-corrected GloFAS and Noah-MP-based historical monthly inflow simulations relative to in situ observations for New Bullards Bar Reservoir (BUL), California. The time series compares observed monthly mean inflow with raw GloFAS, bias-corrected GloFAS, and Noah-MP simulations during 2007–2023. The scatterplot compares monthly inflow simulated by bias-corrected GloFAS and Noah-MP against in situ observations. The dashed diagonal line represents the 1:1 relationship.



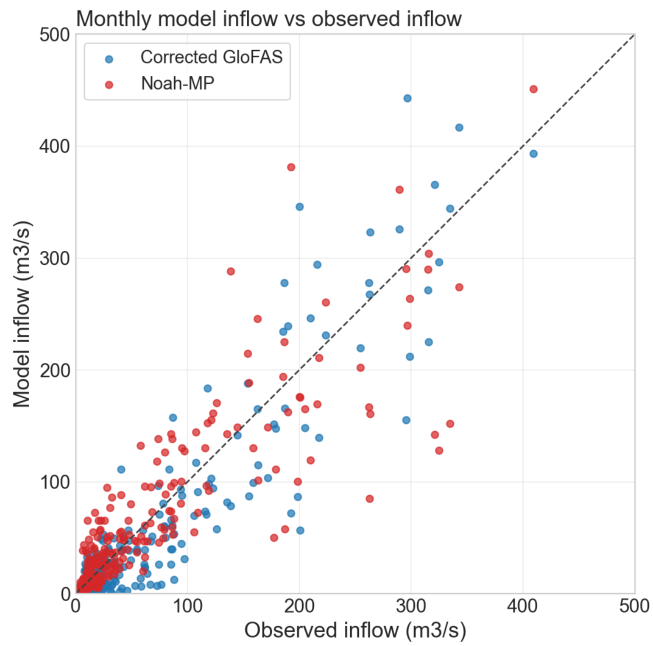


Figure B2. Evaluation of bias-corrected GloFAS and Noah-MP-based historical monthly inflow simulations relative to in situ observations for Don Pedro Reservoir (DNP).

Section S2. Evaluation of Inflow Forecasts Initialized during the Low-Flow Period

Figures B3 and S4 compare monthly Noah-MP-based Ensemble Streamflow Prediction (ESP) and GloFAS inflow forecasts initialized from May through August in 2024 and 2025, with forecast leads extending from each initialization date through the end of the corresponding water year, at New Bullards Bar Reservoir (BUL) and Don Pedro Reservoir (DNP), respectively. As shown in Figures B3 and S4, both forecast methods capture the seasonal decline in reservoir inflow from late spring toward the summer low-flow period. Forecasts initialized in May show the largest inflow magnitudes and the widest ESP ensemble spread, particularly at BUL, whereas forecasts initialized in July and August remain consistently low throughout the forecast period. BUL generally exhibits larger inflow magnitudes and greater forecast uncertainty than DNP, while inflows at DNP are lower and the forecast means remain relatively close to the observations during most low-flow months. At both reservoirs, ESP and GloFAS reproduce the broad transition toward low-flow conditions, although differences are evident during the early forecast months and around individual inflow peaks. The ESP ensemble spread generally increases with lead time, especially near the end of the forecast horizon. Overall, the results indicate that both ESP and GloFAS reasonably capture the seasonal evolution of low-flow inflows, while uncertainty is greater for earlier initialization dates, longer lead times, and the higher-flow conditions at BUL.

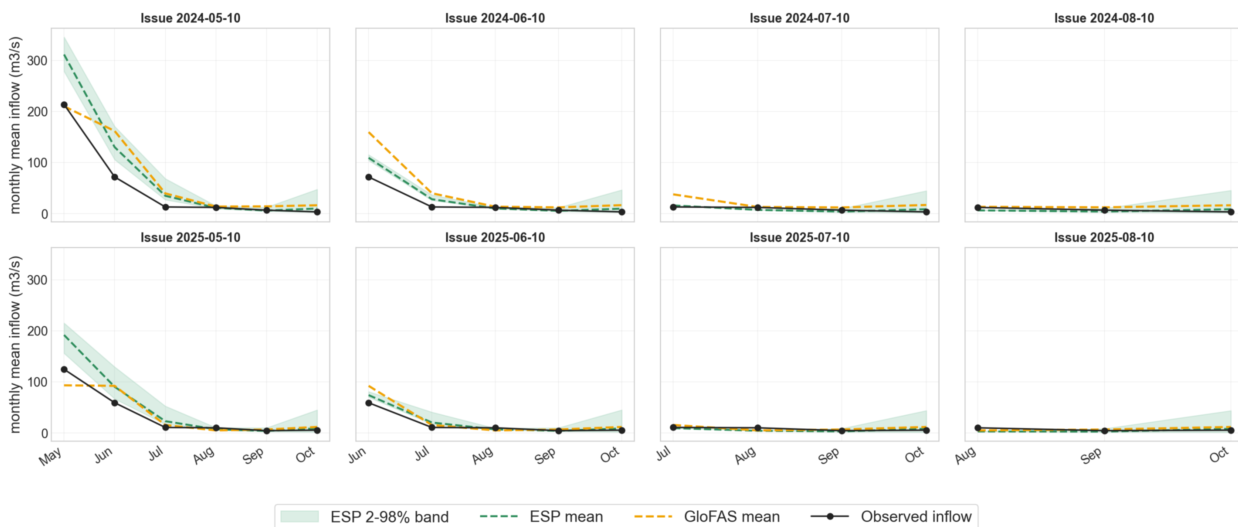


Figure B3. Comparison of ESP and GloFAS monthly inflow forecasts initialized during the May–August low-flow period at New Bullards Bar Reservoir (BUL) for 2024 and 2025.

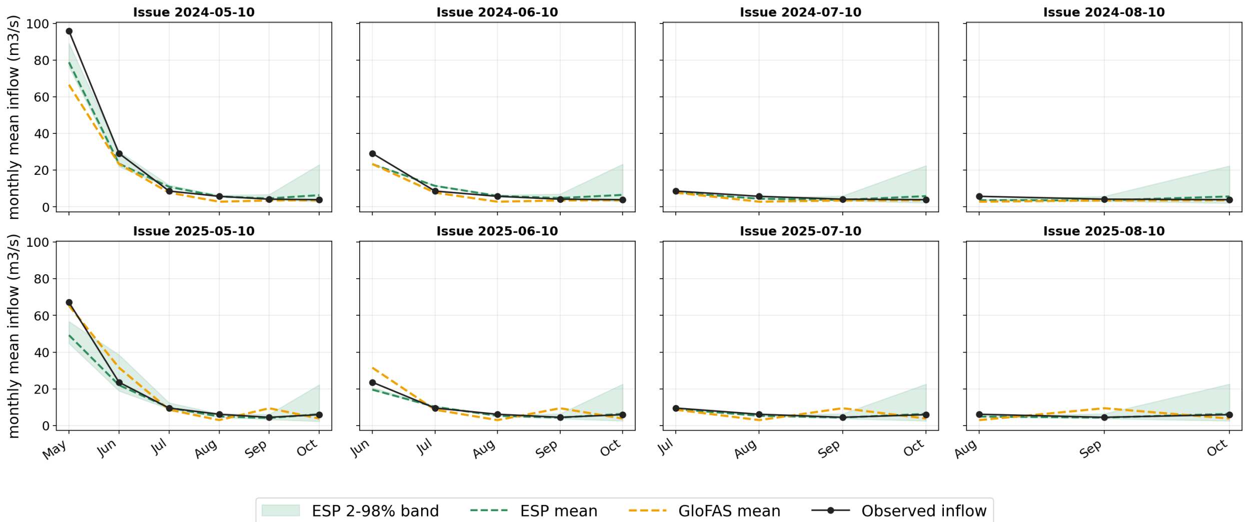


Figure B4. Comparison of ESP and GloFAS monthly inflow forecasts initialized during the May–August low-flow period at Don Pedro (DNP) for 2024 and 2025.

Section S3. SWOT-observed Water Surface Elevation and Derived Storage at Tendaho Reservoir

Figure B5 shows the time series of SWOT-observed water surface elevation (WSE) and corresponding reservoir storage at Tendaho Reservoir (TDH) from early 2024 to mid-2026. Storage was derived from SWOT WSE and surface-area observations using the estimation method described in Section 4.3 of the main text. Because reservoir storage was derived directly from SWOT WSE and surface-area observations, the WSE and storage time series exhibit closely aligned temporal variations (Figure B5). Both variables show a pronounced seasonal cycle, with reservoir filling during mid- to late 2024 and again in late 2025, followed by gradual drawdown during the subsequent dry periods. The highest observed WSE and storage occurred in late 2024, whereas lower reservoir levels were observed in mid-2025 and again toward mid-2026. These observations characterize the seasonal storage states of TDH and provide the reservoir initial conditions used in the subsequent storage forecasting experiments in data-limited regions.

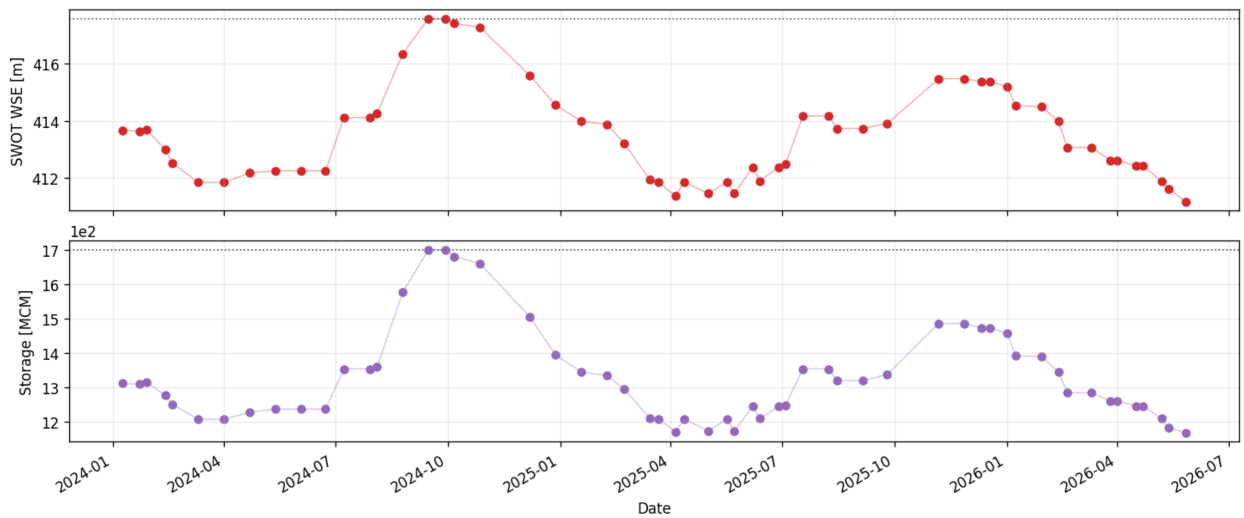


Figure B5. Time series of SWOT-observed water surface elevation and derived storage at Tendaho Reservoir (TDH), Ethiopia.