Reproducibility of Daily Unimpaired Flow Computations:
Case Study of San Joaquin Basin Tributaries, California

By

YARA MARTIN PASNER
THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Hydrologic Sciences

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Jay R. Lund
Jay R. Lund, Chair

Jonathan D. Herman

Samuel Sandoval Solis

Committee in Charge

2021

ABSTRACT

Monthly unimpaired flow calculations have been at the heart of hydrologic forecasting and water management in California for decades. Yet only recently have daily unimpaired flow calculations been applied directly to state-level water management policy. In 2018, the California State Water Resources Control Board (SWRCB) resolved to incorporate daily unimpaired flow calculations into their ongoing effort to support threatened native fish populations in the San Francisco Bay/Sacramento-San Joaquin Delta Estuary. For major tributaries to the San Joaquin Delta (the Stanislaus, Merced, and Tuolumne rivers), instream flow requirements are set to be a fixed percentage of the daily unimpaired flow, averaged over a seven-day rolling window. This novel and contentious application of daily unimpaired flow values to instream flow requirements motivated the SWRCB to investigate the calculation of daily unimpaired flow values for the Stanislaus, Merced, and Tuolumne rivers. This investigation found that daily unimpaired flow values are computed in real time using provisional data, which makes it difficult to reproduce historical daily unimpaired flow values with perfect reliability. Lack of reliable daily unimpaired flow computation reproducibility is potentially problematic for the SWRCB as their updated instream flow requirement directly applies these computations to the allocation of water resources between regional water agencies, farmers, and fish. Therefore, discrepancies between reproduced daily unimpaired flows values and those reported online might fuel existing litigative action against the SWRCB. To support the SWRCB, a quantitative framework is proposed to assess trends in unimpaired flow reproducibility. This framework is illustrated by reproducing 11 years of historical daily unimpaired flow calculations for the Stanislaus, Merced, and Tuolumne rivers. The results of this case study are used to discuss potential causes of discrepancies between reproduced and reported daily unimpaired flow values, and show that for each computation, daily unimpaired flow values are less reproducible when flows are low (during extreme drought). Further research on the precise cause of discrepancies could illuminate potential impacts of computation reproducibility on the SWRCB's updated instream flow requirement, and on hydrologic modeling in general.

ACKNOWLEDGEMENTS

First, I would like to thank Professor Jay Lund, whose feedback and support were invaluable in formulating the analysis presented in this thesis. I also thank Professors Sam Sandoval Solis and Jon Herman for recommending goodness-of-fit measures and computing tools that were integral to the methodology of this thesis, and for serving on my thesis committee. And I thank Amber Pulido, Lindsay Murdoch, and Bill Fleenor for your hard work and collaboration throughout this research project.

Next, I acknowledge and thank the SWRCB for providing funding to support my time as an M.S. student, and for leading the state-wide endeavor to support native fish species that are threatened by human alterations to California's hydrologic system – may this effort be successful. In particular, I thank the following staff members at the SWRCB for contributing to this report: Erin Foresman, Nicole Williamson, Reza Ghasemizadeh, and Yongxuan Gao. Additionally, I thank Wes Monier and members of the staff at Turlock ID for expert and earnest cooperation with this study. And I thank staff at the California Department of Water Resources for insight on the computation and curation of daily unimpaired flow values, including Ashok Bathulla, Sean DeGuzman, and Steve Nemeth (retired).

On a personal level, I thank Justin Woodjack for support and encouragement throughout my time as a master's student and beyond as we cope with historic times of quarantine and isolation in a world ravaged by COVID-19. I also thank my brother, Jake Pasner, for debating the architecture of my code with me, and for taking care of the farm with Dad while I studied. And finally, I thank my parents, Izzy Martin and Mike Pasner, for my overall existence, and for the lifetime of personal and financial support that has made this degree possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1. INTRODUCTION AND OBJECTIVES

1.1. Introduction to Unimpaired Flow

Understanding how precipitation and snowpack relate to runoff is essential to accurate hydrologic forecasting. Precipitation and snowpack have a natural relationship with runoff, but in California, this natural relationship is affected and often overpowered by extensive human alterations to the surface water and groundwater management system. Therefore, hydrologic forecasters often use unimpaired flow values in place of observed runoff to model the relationship between runoff and precipitation.

Although unimpaired flow data are computed in a decentralized way by water managers throughout the state, the California Department of Water Resources (DWR) oversees a more systematic curation of unimpaired flow values. The California Department of Water Resources Bay Delta Office (DWR-BDO) defines unimpaired flow as a "…theoretically available water supply assuming existing river channel conditions in the absence of (1) storage regulation for water supply and hydropower purposes and (2) stream diversions for agricultural and municipal uses" (DWR-BDO, 2016). In 2000, the California Department of Water Resources Division of Flood Management (DWR-DFM) released a memorandum report titled the "Derivation of Unimpaired Runoff in the Cooperative Snow Surveys Program." This Unimpaired Runoff Memorandum describes how monthly unimpaired flow values are computed for each of the major streams in California for which forecasts are made by the California Cooperative Snow Surveys Program. This report was revised in 2016 (DWR-DFM, 2016).

According to the revised Unimpaired Runoff Memorandum (DWR-DFM, 2016), monthly unimpaired flow values are computed using a mass balance approach: upstream diversions, exports, reservoir storage change, and reservoir evaporation values are added to the observed runoff at a given location, while irrigation return flows and imports are subtracted from the flow (DWR-DFM, 2016). Each unimpaired flow value is computed by a designated local, state, or federal water agency, referred to in this report as the "computing agency." Once unimpaired flow values are computed for a given location, computing agencies report these values onto DWR's California Data Exchange Center (CDEC). On CDEC, unimpaired flow values are referred to as "full natural flows" (FNF). This terminology is used interchangeably with "unimpaired flows" throughout this study. Some computing agencies report daily FNF values to CDEC every working day, while others only report monthly FNF values. This thesis focuses only on daily FNF values because these values have recently been applied directly to state-level water management policy for the first time.

1.2. Background

In addition to use for hydrologic forecasting, unimpaired flow values inform water management and policy throughout California. Since 2006, monthly FNF values posted to CDEC have been used to inform minimum instream flow requirements for tributaries to the San Francisco Bay/Sacramento-San Joaquin Delta Estuary (the "Bay-Delta") under the SWRCB Water Quality Control Plan (SWRCB, 2006). In 2018, the SWRCB updated their Water Quality Control Plan (SWRCB, 2018). This update refined the instream flow objective for major tributaries to the Bay-Delta from monthly to daily minimum flow requirements. In the updated plan, daily instream flow requirements were defined to be a fixed percentage of the rolling seven-day average of computed daily FNF value posted to CDEC (SWRCB, 2018).

This application of daily FNF to minimum instream flow requirements is perhaps the most direct, contentious use of daily FNF values in California water management to date. Because of this novelty, in 2019 the SWRCB initiated a detailed investigation of processes that produce daily FNF values for the three major tributaries to the Lower San Joaquin River: the Stanislaus, Merced, and Tuolumne rivers (Pulido *et al.* 2021). This investigation examined the mass balance equations for the daily unimpaired flow of each of these three tributaries, which differ slightly from the monthly equations defined by the 2016 revised Unimpaired Runoff Memorandum report. Additionally, Pulido *et al.* found that for these three rivers, daily FNF values are computed in real time using provisional data, making it difficult to reliably reproduce historical daily FNF values posted to CDEC (Pulido *et al.* 2021). To evaluate how this lack of reliable daily FNF reproducibility might impact the SWRCB's daily instream flow requirements, a framework to quantify trends in the reproducibility of daily FNF computations should be established.

1.3. Thesis Objective and Structure

This thesis proposes a quantitative framework for assessing the reproducibility of daily FNF values posted on CDEC. This framework is illustrated by reproducing 11 years of historic daily FNF values for the three major tributaries to the Lower San Joaquin River: the Stanislaus, Tuolumne, and Merced rivers. Using this framework, seasonal and multi-year trends in reproducibility are identified for each river. These findings are presented in the results section. Potential modifications to this method of analysis are presented in the discussion, along with possible causes of discrepancies between reproduced and reported daily FNF values. Recommendations for additional research on methods used to reproduce daily FNF values, on the efficacy of this framework of analysis, and on the physical implications of daily FNF reproducibility are discussed in the future work section.

2. METHODS

This section begins by introducing the case study area and analysis period. This is followed by a detailed description of the methods used to reproduce 11 years of historic daily FNF values for the Stanislaus, Tuolumne, and Merced rivers. Next, in Section 2.7, the framework established to analyze how well reproduced daily FNF values match values posted on CDEC is described. This section closes with a summary of procedures used to reproduce and analyze daily FNF values in this case study.

2.1. Study Area and Analysis Period

To illustrate our proposed framework of analyzing reproducibility of daily FNF values, 11 years of historic daily FNF values were reproduced for three unimpaired flow gauges on major tributaries to the Lower San Joaquin River (Figure 1). These gauges are for Stanislaus River at Goodwin Dam, Tuolumne River below La Grange Dam, and Merced River below Merced Falls Dam. The river basin, CDEC Station ID, and computing agency for each gauge are reported by Table 1. Throughout this report, unimpaired flow gauges are referred to by their river basin for simplicity.

The Stanislaus, Tuolumne, and Merced rivers were chosen for study to support the implementation of the SWRCB's 2018 Water Quality Control Plan. Daily FNF values for these three rivers are reproduced for water years (WY) 2009-2019. This period was chosen because it includes both an extreme drought (WY 2012-2016) and an extremely wet year (WY 2017).

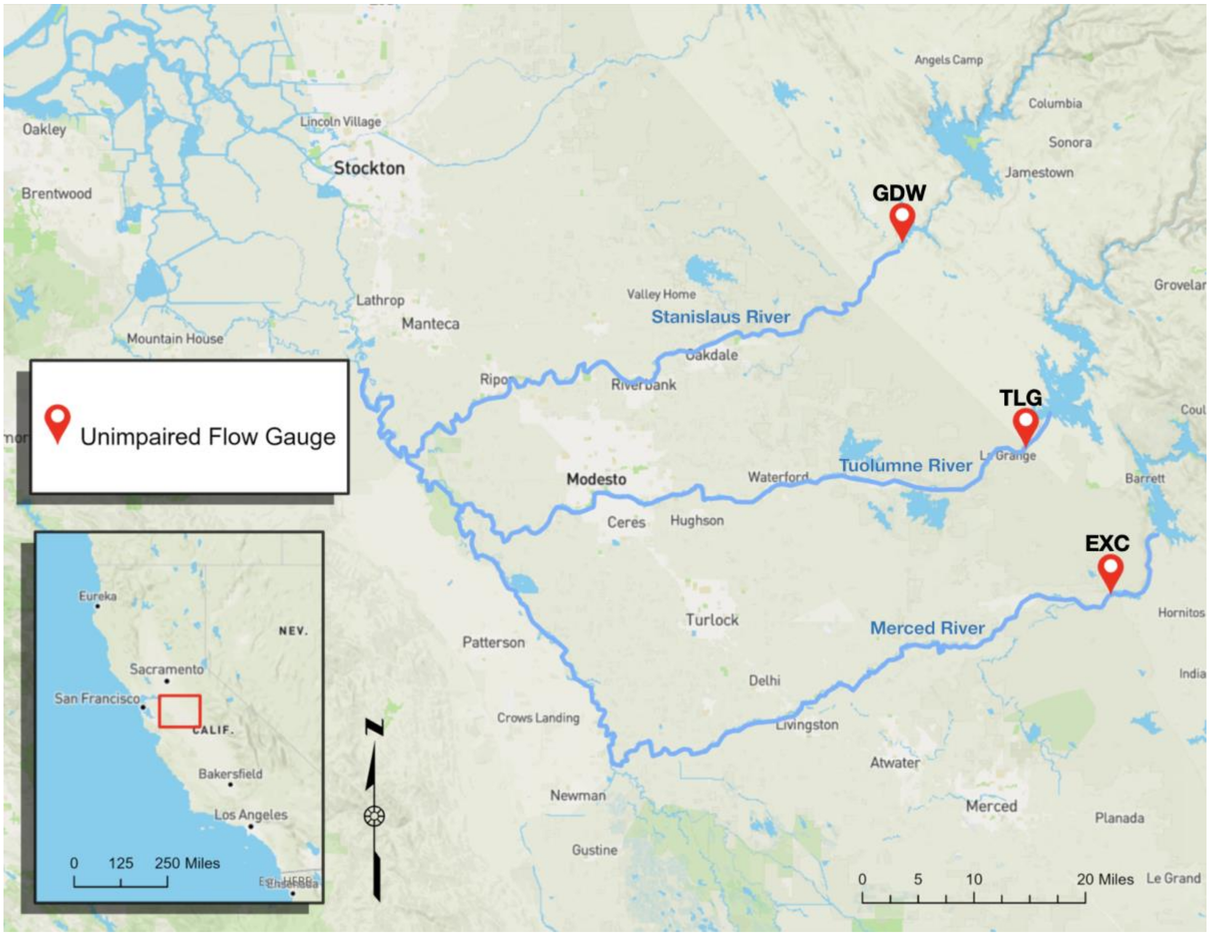


Figure 1: Unimpaired Flow Gauges for Major Tributaries to Lower San Joaquin River, California

Table 1: Summary of Unimpaired Flow Gauges Considered in Study

Unimpaired Flow Gauge	River Basin	CDEC Station ID	Computing Agency
Stanislaus River at Goodwin Dam	Stanislaus	GDW	DWR-DFM
Tuolumne River below La Grange Dam	Tuolumne	TLG	Turlock Irrigation District (Turlock ID)
Merced River below Merced Falls Dam	Merced	EXC	Merced Irrigation District (Merced ID)

2.2. Daily Full Natural Flow Equations

Although *monthly* FNF equations for the Stanislaus, Tuolumne, and Merced rivers are published by DWR-DFM in the Unimpaired Runoff Memorandum (2016), equations used to compute *daily* FNF are not available to the public. Previous research has shown that these daily FNF equations are based on the associated monthly FNF equation, but some impairment terms are excluded from the daily FNF equation (Pulido *et al.* 2021).

In this case study, Equations 1 through 3 are used to compute daily FNF for the Stanislaus River at Goodwin Dam, Tuolumne River below La Grange Dam, and Merced River below Merced Falls Dam respectively. These equations are provided by the Unimpaired Flow Technical Report by Pulido *et al.* (2021). Each equation was either confirmed by the daily FNF computing agency or verified empirically using the best available data (Pulido *et al.* 2021).

Equation 1: Daily FNF Equation for Stanislaus River at Goodwin Dam

Daily FNF (cfs) = **Measured Gage Flow** (Stanislaus River at Goodwin Dam)
+ **Exports** (Goodwin N Main Canal + Goodwin S Main Canal + Farmington Central ID Canal + Farmington Stockton E Canal)
+ **Diversion** (Tuolumne Canal)
+ **Evaporation** (New Melones Reservoir)
+***Δ*Storage** (New Melones + Spicer Meadows + Beardsley Lake + Donnells + Tulloch + Strawberry + Relief + Lyons)

Equation 2: Daily FNF Equation for Tuolumne River below La Grange Dam

Daily FNF (cfs) = **Measured Gage Flow** (Tuolumne River below La Grange Dam)
+ ***Δ*Storage** (Don Pedro + Lake Eleanor + Cherry Valley + Hetch Hetchy)
+ **Evaporation** (Don Pedro)
+ **Diversion** (Diversion to S.F. Pipeline + Turlock Canal + Modesto Canal)

Equation 3: Daily FNF Equation for Merced River below Merced Falls Dam

Daily FNF (cfs) = **Measured Gage Flow** (Merced River below Merced Falls Dam)
+ ***Δ*Storage** (Lake McClure 'Exchequer' + Lake McSwain)

2.3. Data Availability

Most data needed to reproduce daily FNF values for the Stanislaus, Tuolumne, and Merced rivers are available online, either through CDEC (link) or the United States Geological Survey's National Water Information System (NWIS). In addition to online data, computing agencies often retain an in-house version of datasets used to compute daily FNF. These three data sources often have differing numerical values, as each institution has their own quality assurance and quality control (QA/QC) procedures for revising provisional data.

This section describes the availability of online and in-house computing agency data for the WY 2009-2019 computation of daily FNF for the Stanislaus, Tuolumne, and Merced rivers. Note that if both storage change (sensor 22) and storage data (sensor 15) were available on CDEC for a given reservoir, sensor 15 data were prioritized. This decision is informed by the analysis presented in Pulido *et al.* 2021, where sensor 15 data was largely found to provide a better reproduction of daily FNF than sensor 22 data.

2.3.1. Stanislaus River

Table 2 lists online data considered for the Stanislaus daily FNF computation. In Table 2, data sources used for this study are highlighted in green, showing that CDEC data were used wherever available except for Tuolumne Canal diversions and for Lyons Reservoir storage. In these two instances, NWIS data were used to reproduce daily FNF. The selection of these sources is discussed in Section 2.6.

Because DWR-DFM is the Stanislaus FNF computing agency, datasets used by DWR-DFM to calculate Stanislaus River daily FNF are all available on CDEC, and no in-house computing agency data are available (or needed).

Table 2: Datasets Available Online for Stanislaus River Daily FNF Calculation (WY 2009-2019)

Name	Term Type	NWIS	CDEC
Stanislaus River at Goodwin Dam	*Measured Gage Flow (cfs)*	11302000	GDW – 71
Goodwin N Main (South San Joaquin) Canal	*Export (cfs)*	11300500	GDJ – 85
Goodwin S Main (Oakdale) Canal	*Export (cfs)*	11301000	GDS – 85
Farmington Central ID Canal	*Export (cfs)*	-	FR1 – 85
Farmington Stockton E Canal	*Export (cfs)*	-	FR2 – 85
Tuolumne Canal near Long Barn	*Diversion (cfs)*	11297500	STU – 110
New Melones Reservoir	*Evaporation (cfs)*	-	NML – 74
New Melones Reservoir	*Storage Change (ac-ft)*	11299000	NML-15
Beardsley Lake	*Storage Change (ac-ft)*	11292800	BRD-15
Donnells Reservoir	*Storage Change (ac-ft)*	11292600	DON-15
Tulloch Reservoir	*Storage Change (ac-ft)*	11299995	TUL-15
Spicer Meadows	*Storage (ac-ft)*	11293770	SPM – 15
Strawberry Reservoir	*Storage (ac-ft)*	-	SWB – 15
Relief Reservoir	*Storage (ac-ft)*	-	RLF – 15
Lyons Reservoir	*Storage (ac-ft)*	11297700	LYS – 15

2.3.2. Tuolumne River

Table 3 lists all online data considered for the daily FNF calculation of Tuolumne River. In Table 3, data sources used for this study are highlighted in green. Table 3 shows that NWIS data were used wherever available except for Don Pedro Reservoir and Lake Eleanor storage, where CDEC data were used. The process of choosing these sources is discussed in Section 2.6.

In addition to online data, the Tuolumne FNF computing agency (Turlock ID) provided a complete record of computing agency data for each Tuolumne FNF calculation component for WY 2009-2019.

Table 3: Datasets Available Online for Tuolumne River Daily FNF Calculation (WY 2009-2019)

Name	**Term Type**	**NWIS**	**CDEC**
Tuolumne River below La Grange Dam	*Measured Gage Flow (cfs)*	11289650	LGN-41
Don Pedro Reservoir	*Storage (ac-ft)*	11287500	DNP – 15
Lake Eleanor Reservoir	*Storage (ac-ft)*	11277500	ENR – 15
Cherry Valley Reservoir	*Storage (ac-ft)*	11277200	CHV – 15
Hetch Hetchy Reservoir	*Storage (ac-ft)*	11275500	HTH – 15
Don Pedro Reservoir	*Evaporation (cfs)*	-	DNP – 74
Modesto Canal near La Grange, CA	*Diversion (cfs)*	11289000	-
Turlock Canal near La Grange, CA	*Diversion (cfs)*	11289500	-

2.3.3. Merced River

Table 4 lists all online data considered for the daily FNF calculation of Merced River. In Table 4, data sources used for this study are highlighted in green, showing that only NWIS data were used in this reproduction. The selection of these sources is discussed in Section 2.6.

For the Merced River, no computing agency data were available for the 11-year period of study. Computing agency data were requested, but Merced ID did not provide these data.

Table 4: Datasets Available Online for Merced River Daily FNF Calculation (WY 2009-2019)

Name	**Term Type**	**NWIS**	**CDEC**
Merced River below Merced Falls Dam	*Measured Gage (cfs)*	11270900	MMF – 20

Lake McClure (Exchequer) Reservoir	*Storage (ac-ft)*	11269500	EXC – 15
Lake McSwain Reservoir	*Storage (ac-ft)*	11270600	MCS - 15

2.4. Data Management

This section describes the methods used in this analysis to manage problematic data, such as data gaps, not-a-number entries, and negative FNF values.

2.4.1. Data Gaps

Any day without a numeric value recorded for a specific dataset is considered a data gap. For data gaps encountered in this study, a generalizable procedure was established to ensure data gaps were filled consistently, in a mass-conservative manner. This procedure is intended to align with the everyday data management practices of FNF computing agencies. However, further research on river-specific data management practices might improve reproducibility of daily FNF computations. The following procedure was used to manage daily FNF input data:

1. Import data into program of choice. This analysis was conducted with Python3, but Microsoft Excel, R, and MATLAB are examples of alternative programs that could be used.
2. Linearly interpolate over small data gaps (30 days or less).
3. Manipulate data as needed for the FNF equation.

    For example…

    - To convert hourly data to a daily timestep, calculate the daily mean from hourly data
    - To compute daily storage change from daily storage data, subtract the storage value on the previous day from the storage value on the date of interest for each day of study
    - To convert daily data from ac-ft to cfs, multiply each value by 0.5042 cfs/ac-ft

4. Fill data gaps larger than 30 days with the mean value of the remaining WY 2009-2019 dataset.

Any day where no daily FNF value was reported to CDEC was dropped from the analysis. Only three days out of 11 years of study did not have daily FNF values reported on CDEC (Figures 3 and 4).

Figures 2 through 4 show data gaps encountered in the WY 2009-2019 period of study. In these figures, each colored diamond represents one day where no numeric value was recorded for the station on the vertical axis. A data gap of multiple consecutive days forms a line (the colors have no particular meaning). No data gaps were encountered in Turlock ID or NWIS input data for the Tuolumne River computation (Figure 3).

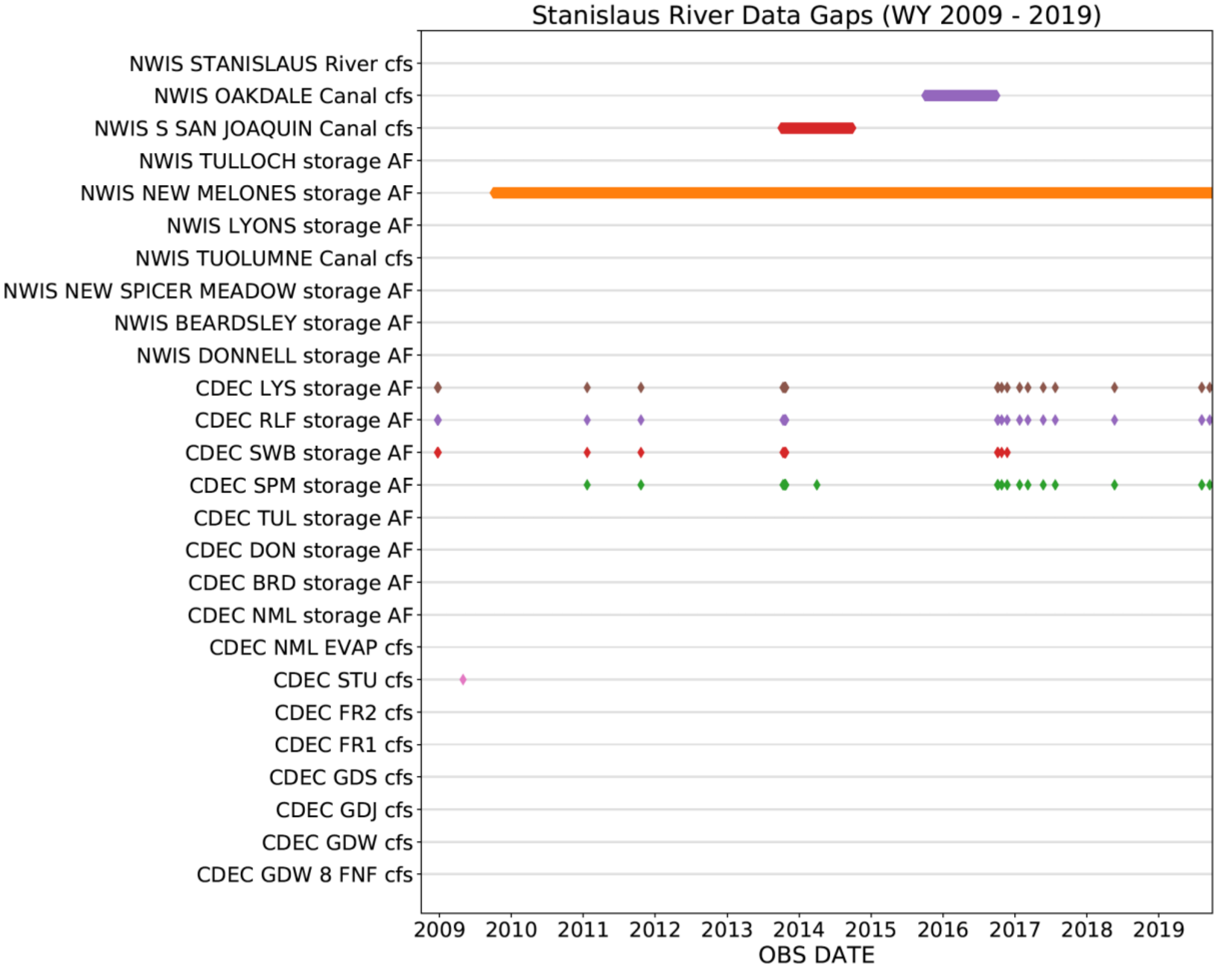


Figure 2: Data Gaps in the Stanislaus River WY 2009-2019 Daily FNF Calculation.

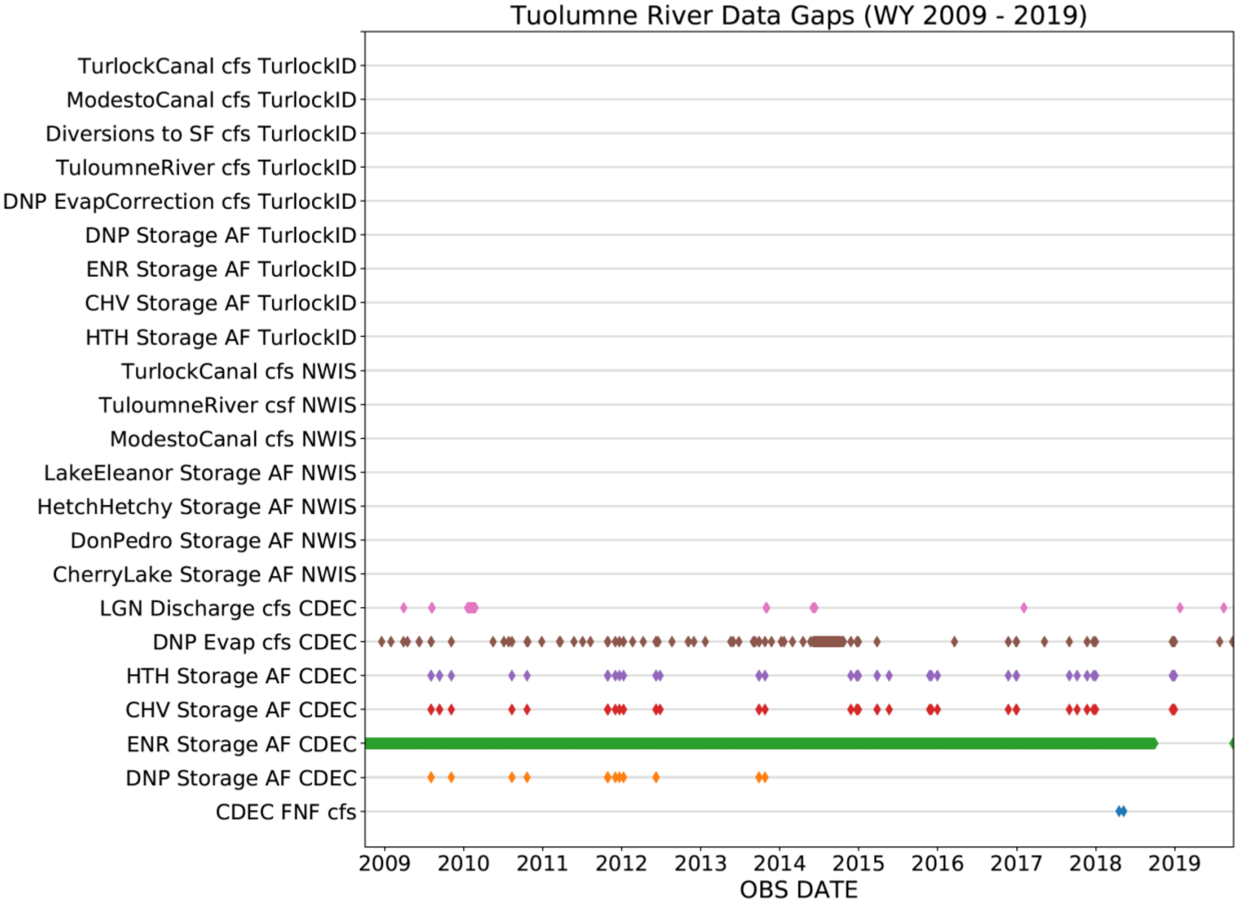


Figure 3: Data Gaps in the Tuolumne River WY 2009-2019 Daily FNF Calculation.

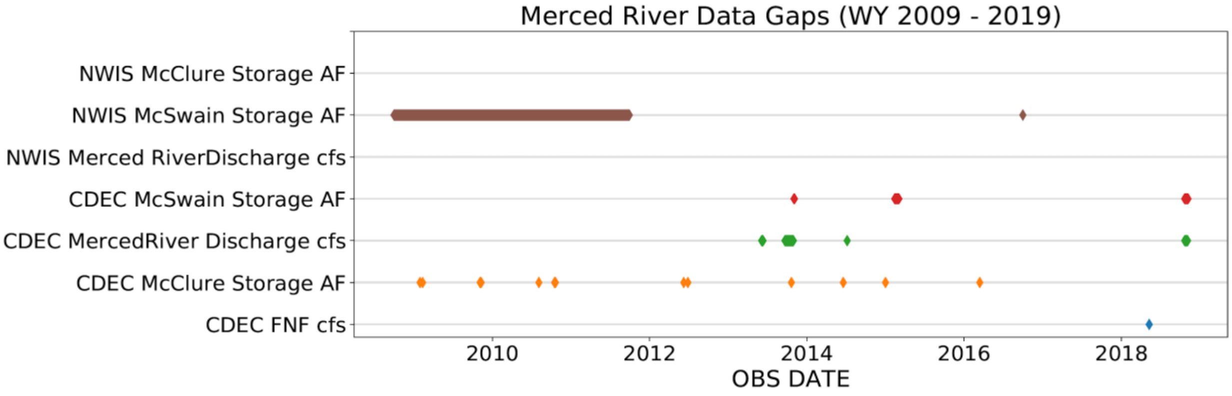


Figure 4: Data Gaps in the Merced River WY 2009-2019 Daily FNF Calculation.

2.4.2. Negative Reproduced FNF Values

In addition to filling data gaps, a secondary data management technique was used after the computation of FNF and before any analysis: All reproduced FNF values computed to be less than zero were replaced with zero. In general, this replacement helped reproduced FNF values align better with FNF values reported on CDEC, although there were some instances where negative FNF values were reported on CDEC. Negative FNF values largely occur when FNF is low, where negative storage change values overwhelm positive terms in the daily FNF equation (such as discharge, evaporation, or diversion terms).

2.5. Computation Procedures

After confirming the daily FNF equations, obtaining all available data, and managing data gaps/negative FNF values, the challenge remains of deciding what datasets to use in reproducing daily FNF values. Because CDEC, NWIS, and FNF computing agencies each have their own QA/QC process to revise provisional data, historic data supplied directly by computing agencies can differ from the same data retrieved from CDEC or NWIS. Furthermore, data on CDEC often differs from the same data pulled from NWIS (this is addressed in Section 2.6).

Following Pulido *et al.* (2021), three computation procedures are used to evaluate how data availability impacts overall reproducibility, and how differences between online and in-house computing agency data effect reproducibility. These computation procedures use the same underlying equations (Equations 1 – 3), but each procedure varies the input datasets used:

- **Procedure 1**: *Online Data Only* – Uses only datasets currently published on NWIS and/or CDEC. Any dataset not available on the NWIS or CDEC websites is treated as zero. If estimates for a component flow were accessible on both NWIS and CDEC but had different numerical values, the data source that provided the best FNF reproduction was identified and used (discussed in Section 2.6).
- **Procedure 2**: *Online Data with Supplemental Computing Agency Data* – Similar to Procedure 1, but includes data supplied by computing agencies for components of the calculation not available on CDEC or NWIS. Procedure 2 is only computed when data are not available on CDEC or NWIS for every component of the daily FNF computation.
- **Procedure 3**: *Computing Agency Data Only* – Uses only datasets supplied by computing agencies. Procedure 3 is only computed where computing agency data are available and have numerical values that differ from those on CDEC or NWIS.

As in Pulido *et al.* 2021, comparing the results between Procedures 1 and 2 provides insight on how online data availability impacts daily FNF reproducibility. Comparing Procedures 2 and 3 provides insight on how differences between online and computing agency data change reproducibility.

As stated in Section 2.3.1, datasets used by DWR-DFM to calculate Stanislaus River daily FNF are all available online, and no in-house computing agency is available. Accordingly, only Procedure 1 (online data only) was reproduced for WY 2009-2019 for the Stanislaus River.

As stated in Section 2.3.2, the Tuolumne FNF computing agency (Turlock ID) provided a complete record of computing agency data for each component of the Tuolumne FNF calculation for WY 2009-

2019. These data often differed numerically from corresponding online data, so all three calculation procedures were reproduced for Tuolumne River using Equation 2. Since the only component of the Tuolumne FNF calculation not available online are diversions to the San Francisco pipeline, the only difference between Procedures 1 and 2 is the inclusion of these diversion data in Procedure 2.

For the Merced River daily FNF calculation, only Procedure 1 was reproduced, using Equation 3, because computing agency data were not available for the 11-year period of study (Section 2.3.3).

2.6. Choosing the Best Data Sources

Tables 2 through 4 show that most input datasets needed to compute daily FNF for the Stanislaus, Tuolumne, and Merced rivers are available on both the CDEC and NWIS websites. The DWR-DFM Unimpaired Runoff Memorandum (2016) does not generally indicate what source to use if both NWIS and CDEC data are available for the same impairment. Furthermore, it cannot be assumed that historical datasets will be identical across platforms.

For example, Figure 5 shows how Donnell Reservoir storage change data varies depending on the platform from which the data were sourced. Here we see that CDEC data have a few extreme outliers that are not included in the NWIS dataset. Similar differences were observed for nearly every dataset of study available on both CDEC and NWIS. Differences between these datasets are likely due to differences in how each agency revises their provisional data.

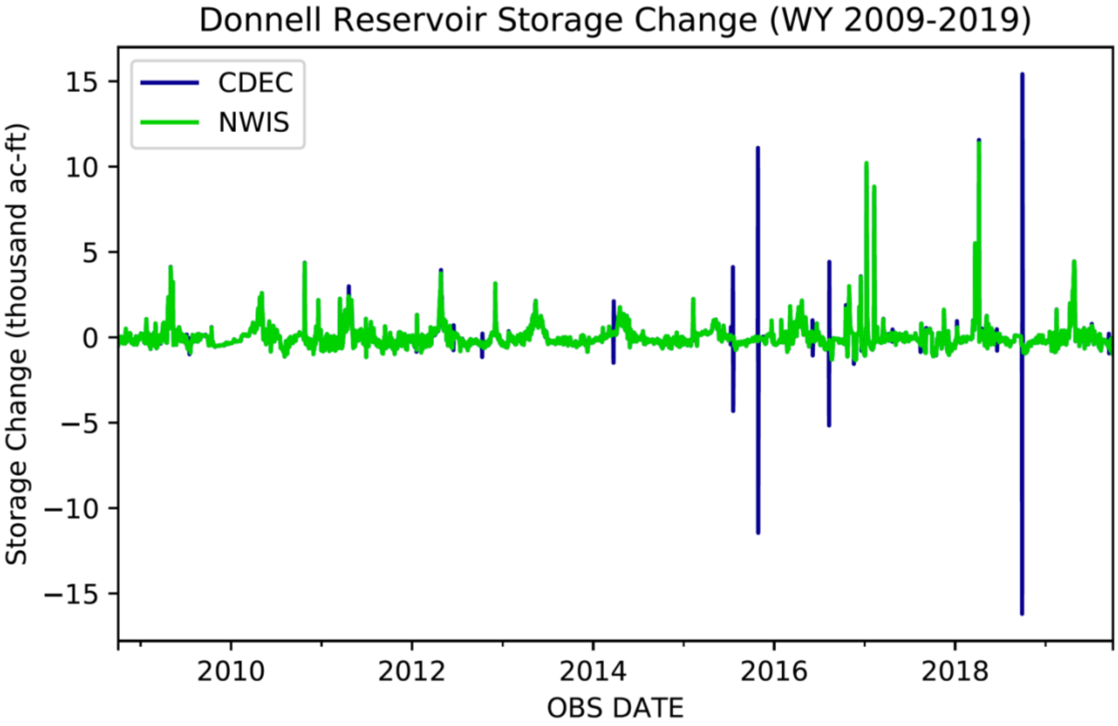


Figure 5: CDEC and NWIS Storage Change Data for Donnell Reservoir, WY 2009-2019

To account for differences between NWIS and CDEC data, the following procedure was developed to determine what combination of online data sources provides the best reproduction of FNF:

1. The Procedure 1 calculation was repeated using every possible combination of online data sources. The number of possible combinations increases exponentially with the number of datasets that are available on both NWIS and CDEC. For example, in the Stanislaus computation there are 10 datasets available on both NWIS and CDEC, resulting in $2^{10} = 1{,}024$ possible combinations of datasets.
2. Each resulting calculation was compared with the daily FNF values posted on CDEC. The goodness-of-fit for each calculation was ranked using the modified Nash-Sutcliffe Coefficient of Efficiency (*NSE*). This choice of metric is discussed in Section 2.7.1.
3. The combination that best reproduces daily FNF from CDEC (having the largest *NSE*) was used for Procedures 1 and 2 in the analyses that follow.

2.7. Evaluating Reproducibility

This section introduces the metrics, tests, and plots that were used to analyze and compare the performance of each daily FNF reproduction throughout WY 2009-2019.

2.7.1. Metrics

Defining metrics to express the reproducibility of an FNF calculation is similar to defining metrics that summarize a hydrologic predictive model's goodness-of-fit. Once a calculation procedure has been reproduced, its output value can be compared to the reported FNF, just as predicted values can be compared against observed. The Nash-Sutcliffe coefficient of efficiency is widely used for this purpose. Researchers and industry leaders commonly employ this metric because of its interpretability and ability to reflect the overall fit of a hydrograph (Moriasi *et al.* 2007).

However, the Nash-Sutcliffe coefficient of efficiency is disproportionately sensitive to high magnitude outliers (Legates and McCabe 1999). This sensitivity is problematic in this analysis because infrequent, high-magnitude discrepancies could misconstrue seasonal or water year trends in overall reproducibility. Therefore, the modified version of the Nash-Sutcliffe coefficient of efficiency defined in Legates and McCabe (1999) is used as the primary metric to analyze reproducibility in the analysis that follows.

The modified Nash-Sutcliffe coefficient of efficiency (*NSE*) is defined by Equation 4, where n is the number of days in the period of study, O_i is the observed value on day i (CDEC FNF), P_i is the predicted value on day i (reproduced FNF), and $\underline{O}$ is the mean value of observed data for all i in range 1 to n.

Equation 4: Modified Nash-Sutcliffe Coefficient of Efficiency (*NSE*)

$$NSE = 1 - \frac{\sum_{i=1}^{n} |O_i - P_i|}{\sum_{i=1}^{n} |O_i - \underline{O}|}$$

NSE ranges from negative infinity to one, where values closer to one indicate better agreement between reproduced and reported daily FNF values. *NSE* can be physically interpreted as the ratio of the mean absolute discrepancy to mean absolute deviation of reported FNF values, subtracted from unity. Here, a discrepancy is defined to be the difference between daily FNF values reported to CDEC and reproduced daily FNF values.

To explore this physical interpretation, consider a time period where the mean absolute daily discrepancy (*MAD*) between reported and reproduced FNF values is larger than the mean absolute deviation of CDEC FNF. Then, the ratio of *MAD* to mean absolute deviation of CDEC FNF would exceed one, and the *NSE* would be negative. Additionally, because the *MAD* exceeds the mean absolute deviation of CDEC FNF, using the average value of CDEC FNF instead of the reproduced timeseries would result in a smaller *MAD* (since in this case, the *MAD* would be equal to the mean absolute deviation of CDEC FNF). Therefore, the mean value of CDEC FNF provides a better reproduction than the reproduced daily FNF timeseries for any computation and time period resulting in a negative *NSE*. Table 5 summarizes this physical interpretation.

Table 5: Physical Interpretation of the Modified Nash-Sutcliffe Coefficient of Efficiency ($\boldsymbol{NSE}$)

	Physical Interpretation
$\boldsymbol{NSE = 1}$	For this period, reproduced FNF values perfectly reproduce the daily FNF values reported on CDEC (*MAD* is equal to zero).
$\boldsymbol{0 < NSE < 1}$	For this period, calculated FNF values do not perfectly reproduce CDEC FNF, but calculated values do perform better than the mean value of CDEC FNF computed over the record of study.
$\boldsymbol{NSE \leq 0}$	For this period, the mean value of CDEC FNF computed over the record of study can serve as a better approximation of daily FNF than reproduced values.

For this study, *NSE* was computed for each daily FNF computation by season (seasonal *NSE*), by water year (water year *NSE*), and for the entire 11-year period of study (11-year *NSE*). The seasonal *NSE* shows reproducibility trends within each water year, while the water year *NSE* shows interannual or multi-year trends. The value of the 11-year *NSE* computed over the entire record of study is used to indicate the overall reproducibility of each FNF calculation. The computation with the highest 11-year *NSE* coefficient is considered the most reproducible, while the computation with the lowest 11-year *NSE* coefficient is considered the least reproducible.

Legates and McCabe (1999) recommend using unitless goodness-of-fit measures like *NSE* in concert with at least one absolute measure that yields results in the units of the variable of study. For this analysis, the mean daily discrepancy is computed by water year and season to measure the average numeric value of discrepancies over time. Additionally, the mean absolute daily discrepancy is used to indicate the average magnitude of daily discrepancies by water year and season. Finally, the ratio of *MAD* to mean CDEC FNF is computed by water year and season to measure the significance of daily discrepancies over time.

2.7.2. Reproducibility Tests

Two forms of reproducibility tests are used for this analysis. The first uses physical reasoning, and the second is a statistical approach. Both tests identify periods when daily FNF computations are less reproducible.

Negativity of NSE

First, if *NSE* is negative for a given computation and time period, using the mean value of CDEC FNF yields a better fit (with a smaller *MAD*) than reproduced FNF values for that time period (Table 5). Using this physical interpretation, we may define the following reproducibility test: if the mean value of CDEC FNF is considered an insufficient representation of daily FNF for a given water year, then any reproduced calculation yielding a negative water year *NSE* value may be considered statistically non-reproducible for that water year.

Similarly, a more stringent version of this reproducibility test was considered by evaluating the seasonal *NSE*: if the mean value of CDEC FNF is considered an insufficient representation of daily FNF for a given season, then any reproduced calculation yielding a negative seasonal *NSE* value may be considered statistically non-reproducible for that season.

Random Permutations Test

The second type of reproducibility test used in this analysis was proposed by Bardsley and Purdie (2007) for predictive hydrologic models. This test proposes a null hypothesis of no calculation reproducibility. This null hypothesis is accepted if there is a sufficiently high probability (greater than five percent) that a random reordering of reproduced daily FNF values would yield a better fit of CDEC FNF than the actual reproduced time series data. If the null hypothesis is not accepted, the test is inconclusive (Bardsley and Purdie 2007).

For this study, a stringent version of Bardsley and Purdie's permutation test was computed for each water year and season using the following procedure. First, testing data are generated for each water year by randomly permuting reproduced daily FNF values intra-seasonally, meaning values cannot permute between years or seasons. This process is repeated to create 100,000 testing datasets for all 11 years of study, providing a test accuracy of two decimal places (Bardsley and Purdie 2007). Seasonal *NSE* and water year *NSE* are then computed for each testing dataset by comparing the randomly permuted reproduced daily FNF values with the unpermuted CDEC FNF for each season and water year of study.

The seasonal test quantity *p* (seasonal *p*-Value) is the number of randomly permuted datasets that yielded an equal or greater seasonal *NSE* coefficient than that of the reproduced FNF values for that season, divided by 100,000. Similarly, the water year test quantity *p* (water year *p*-Value) is the number of randomly permuted datasets that yielded an equal or higher water year *NSE* coefficient than that of the reproduced FNF values for that water year, divided by 100,000.

Both seasonal and water year *p*-Values can be interpreted in the same manner: if five percent or less of permuted datasets matched or outperformed reproduced values of a given season or water year, then the *p*-Value is less than or equal to 0.05, and the test is inconclusive. However, if more than five percent of permuted datasets match or outperform reproduced values for a given water year or season, then the *p*-Value exceeds 0.05, and we accept the null hypothesis of no computation reproducibility for that time period. If the *p*-Value exceeds 0.05 for a water year or season, that computation is considered statistically non-reproducible for that period.

This test can only identify time periods in which a daily FNF computation is statistically non-reproducible. This test cannot show that a computation is statistically reproducible, even if the computation receives the best possible score of $p = 0$ (which indicates that no permuted dataset performed as well as the reproduced data).

2.7.3. Plots

Each of the above metrics and tests were computed for each FNF computation using Python 3. The results were plotted and compared between seasons, water years, river basins, and computation procedures. Additionally, timeseries plots of reproduced FNF vs. CDEC FNF and of daily discrepancies are provided for each computation procedure. In the discrepancy timeseries plot, positive daily discrepancies indicate the reproduced FNF value was an underestimation of CDEC FNF, while negative discrepancies indicate the reproduced FNF value was larger than the associated CDEC FNF value.

2.8. Generalized Procedure for Reproducing Daily FNF

This section provides a generalized summary of the procedures used for this WY 2009-2019 daily FNF case study.

Step 1: Confirm the Daily FNF Equation

Three ways the daily FNF equation can be confirmed are as follows:

1. Confirm the formula with the California Cooperative Snow Surveys Program
2. Confirm the formula with the computing agency responsible for the FNF computation
3. Select the equation and data sources that provide the closest agreement to published CDEC FNF

Step 2: Procure All Available Data

Acquire all relevant data for the period of study in machine-readable format with daily (or finer) temporal resolution, considering the following sources:

1. Available online public data (i.e., NWIS and CDEC)
2. Available computing agency data

Step 3: Format Data for Analysis

1. Import data using program of choice (this analysis was conducted with Python3, but Microsoft Excel, R, and Matlab are alternative programs that could be used)
2. Linearly interpolate over small data gaps (30 days or less)
3. Manipulate data as needed for the FNF equation:
    - To downscale hourly data to a daily timestep, resample data using the daily mean

  - To compute daily storage change from daily storage data, subtract the storage value on the previous day from the storage value on the date of interest for each day of study
  - To convert daily data from ac-ft to cfs, multiply each value by 0.5042 cfs/ac-ft

4. Fill data gaps larger than 30 days with the mean value of the dataset from WY 2009-2019

Step 4: Identify the Best Sources

Compute *NSE* (or some other goodness-of-fit performance metric) for every possible combination of online sources where more than one exist (for example, when CDEC and NWIS present data with different numerical values for the same impairment). A Python3 function was created for this analysis.

Step 5: Calculate Daily FNF

Use the daily FNF equation (from Step 1) and the combination of data sources with the best fit (from Step 4) to compute FNF:

- If no computing agency data were available, compute only Procedure 1
- If all relevant data are available online and computing agency data are also available, compute only Procedures 1 and 3
- If at least one component of the daily FNF computation is not available online, and computing agency data are available, compute Procedures 1, 2, and 3

Finally, for each computation replace any negative daily FNF values with zero.

Step 6: Compute and Compare

Compute the following metrics and tests for each computation of study:

Performance Metrics

- *NSE* by season, water year, and entire record of study
- *MAD* by season and water year
- Ratio of *MAD* to mean FNF by season and water year
- Mean discrepancy by season and water year

Reproducibility Tests

- Non-negativity of seasonal *NSE*
- Non-negativity of water year *NSE*
- Seasonal random permutations test
- Water year random permutations test

Finally, compare the results between water years, seasons, rivers, and computation procedures.

3. RESULTS

The main finding of this case study is that the Stanislaus, Tuolumne, and Merced daily FNF calculations are largely reproducible between WY 2009-2019. However, each computation fails to be statistically reproducible during extremely low-flow periods. This section shows seasonal and water year trends in reproducibility as well as additional computation-specific insights.

3.1. Seasonal and Annual Trends

In general, the Stanislaus is the most reproducible daily FNF calculation, having the highest 11-year *NSE* coefficient for WY 2009-2019 (Table 6). By this same metric, the next most reproducible calculation is Tuolumne Procedure 3 and Merced (tied), followed by Tuolumne Procedure 2, and finally, Tuolumne Procedure 1 (Table 6). Table 6 also shows that the computation with the highest 11-year *NSE* (the Stanislaus) also has the smallest *MAD* and standard deviation of daily discrepancies. Additionally, the computation with the lowest 11-year *NSE* (Tuolumne Procedure 1) has the largest *MAD* and standard deviation of daily discrepancies**.**

Table 6: Summary Statistics of Daily FNF Reproduction, WY 2009-2019

Daily FNF Computation	11-year *NSE*	Mean Discrepancy (cfs)	Mean Absolute Discrepancy (cfs)	Standard Deviation of Daily Discrepancies (cfs)
Stanislaus River	0.957	-11.20	73.38	272.25
Merced River	0.870	-17.40	186.45	465.42
Tuolumne River, Procedure 1	0.830	220.14	466.32	1,072.14
Tuolumne River, Procedure 2	0.857	-9.81	392.65	1,067.87
Tuolumne River, Procedure 3	0.870	-0.708	356.10	835.94

When *NSE* is evaluated by water year, every calculation receives a positive *NSE* value (Figure 6). Therefore, if negativity of water year *NSE* is used to indicate reproducibility, none of the daily FNF calculations are found to be statistically non-reproducible. Similarly, when the water year permutation test is applied, every computation receives *p*-Values equal to zero (the best possible score), indicating the water year permutation test failed to identify any statistically non-reproducible daily FNF calculations.

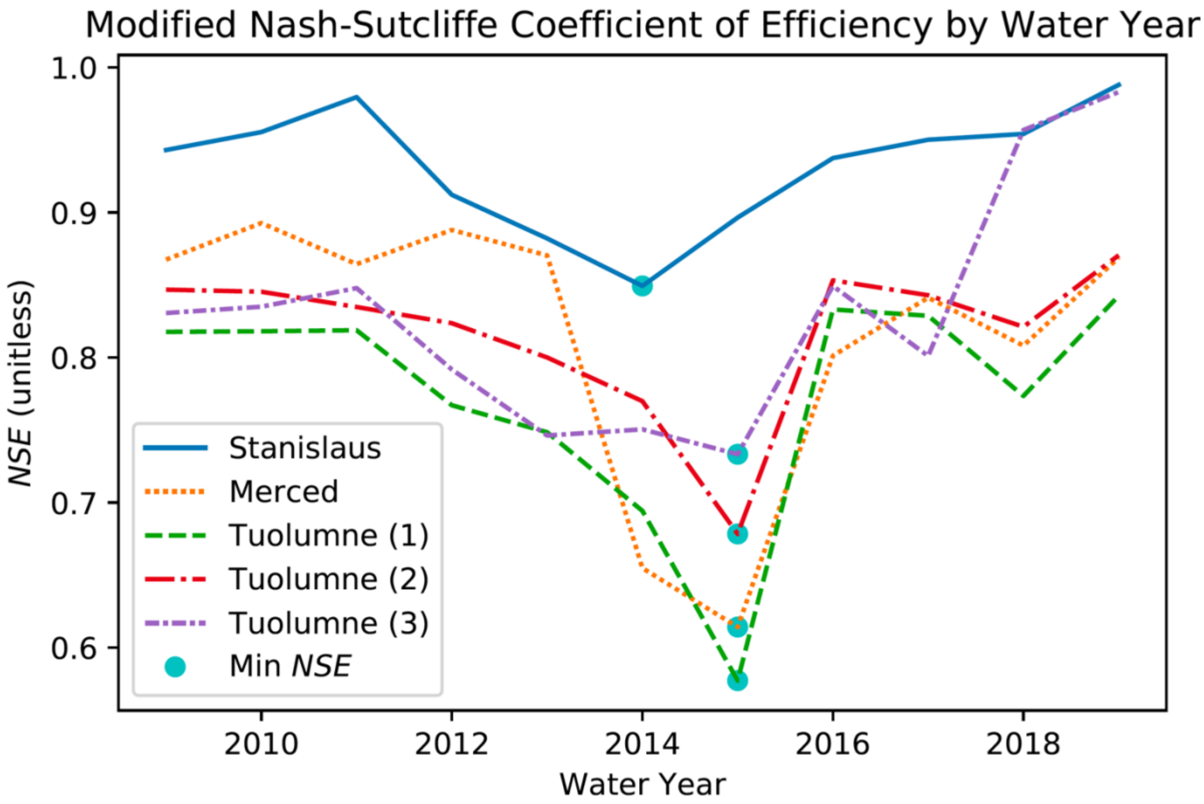
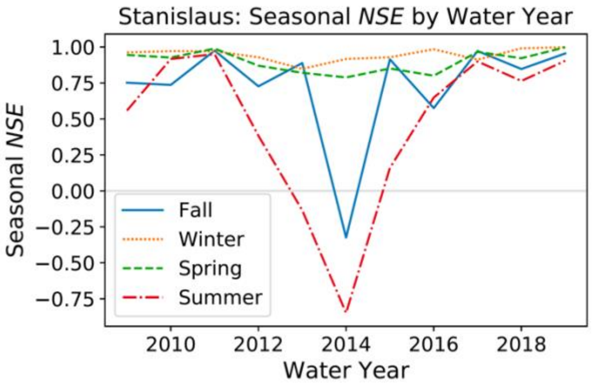
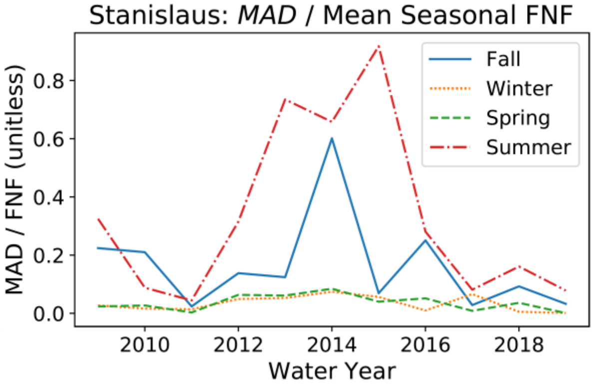
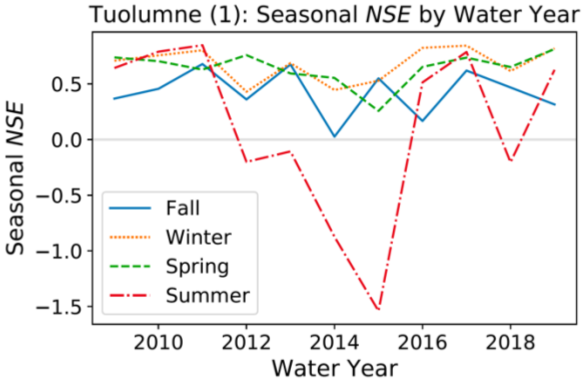


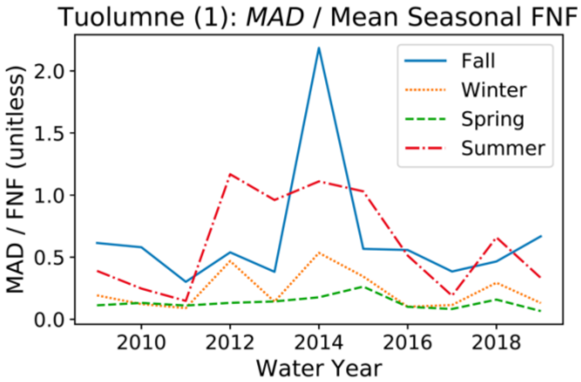
Figure 6: *NSE* by Water Year for Each Daily FNF Calculation, WY 2009-2019

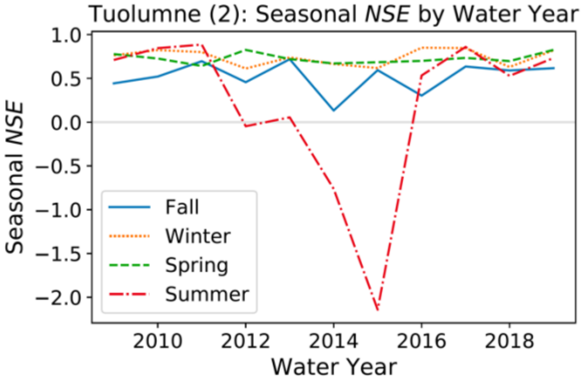
Figure 6 also shows the minimum water year *NSE* value is achieved by each calculation during the WY 2013-2015 drought years. Figure 7 shows that each of these minimum water year *NSE* values correlate with negative seasonal *NSE* values, and highly significant discrepancies (indicated by the ratio of *MAD* to mean FNF) in summer and/or fall of the WY 2013-2015 drought years. If negativity of seasonal *NSE* (more stringent than water year *NSE*) is used to indicate reproducibility, each computation fails to be statistically reproducible for at least one season during the WY 2013-2015 drought years (Figure 7).

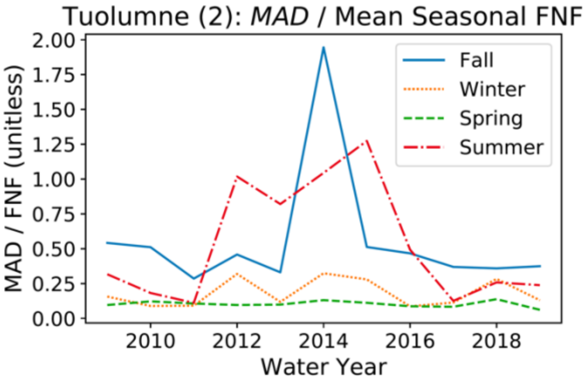

Stanislaus: Seasonal *NSE* by Water Year

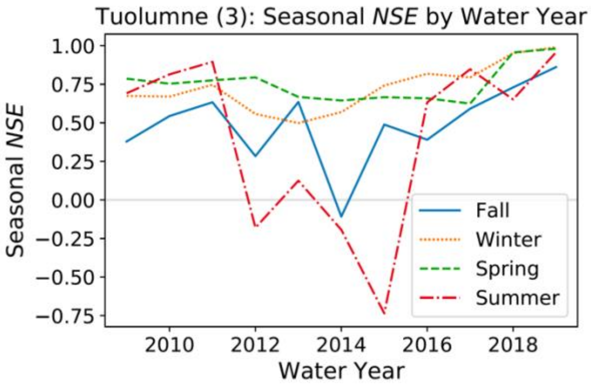

Stanislaus: *MAD* / Mean Seasonal FNF

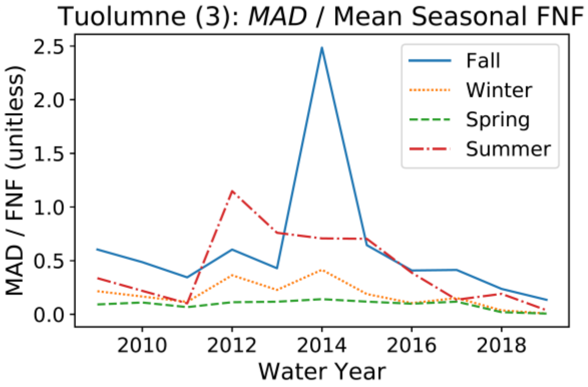

Tuolumne (1): Seasonal *NSE* by Water Year


Tuolumne (1): *MAD* / Mean Seasonal FNF


Tuolumne (2): Seasonal *NSE* by Water Year


Tuolumne (2): *MAD* / Mean Seasonal FNF


Tuolumne (3): Seasonal *NSE* by Water Year


Tuolumne (3): *MAD* / Mean Seasonal FNF

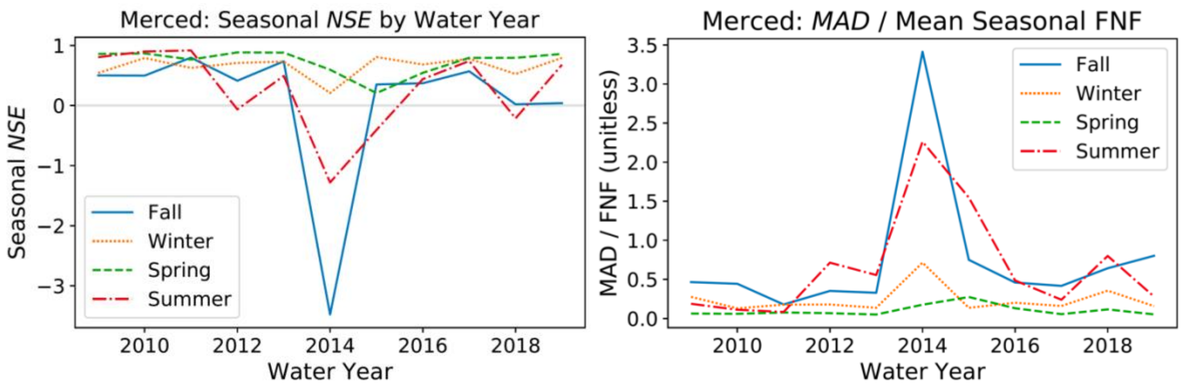


Figure 7: Seasonal *NSE* (left) and the Ratio of Seasonal Mean Absolute Daily Discrepancy to Mean Seasonal FNF (right) by Water Year for Each FNF Calculation, WY 2009-2019

The relationship between flow and reproducibility can be generalized using Figure 8: daily FNF is less reproducible when flows are low, and more reproducible when flows are high. In particular, no daily FNF computation received a negative seasonal *NSE* value during a season where the mean FNF exceeded 600 cfs (Figure 8). And every computation received seasonal *NSE* values above 0.5 during seasons where mean FNF exceeded 1,800 cfs (Figure 8).

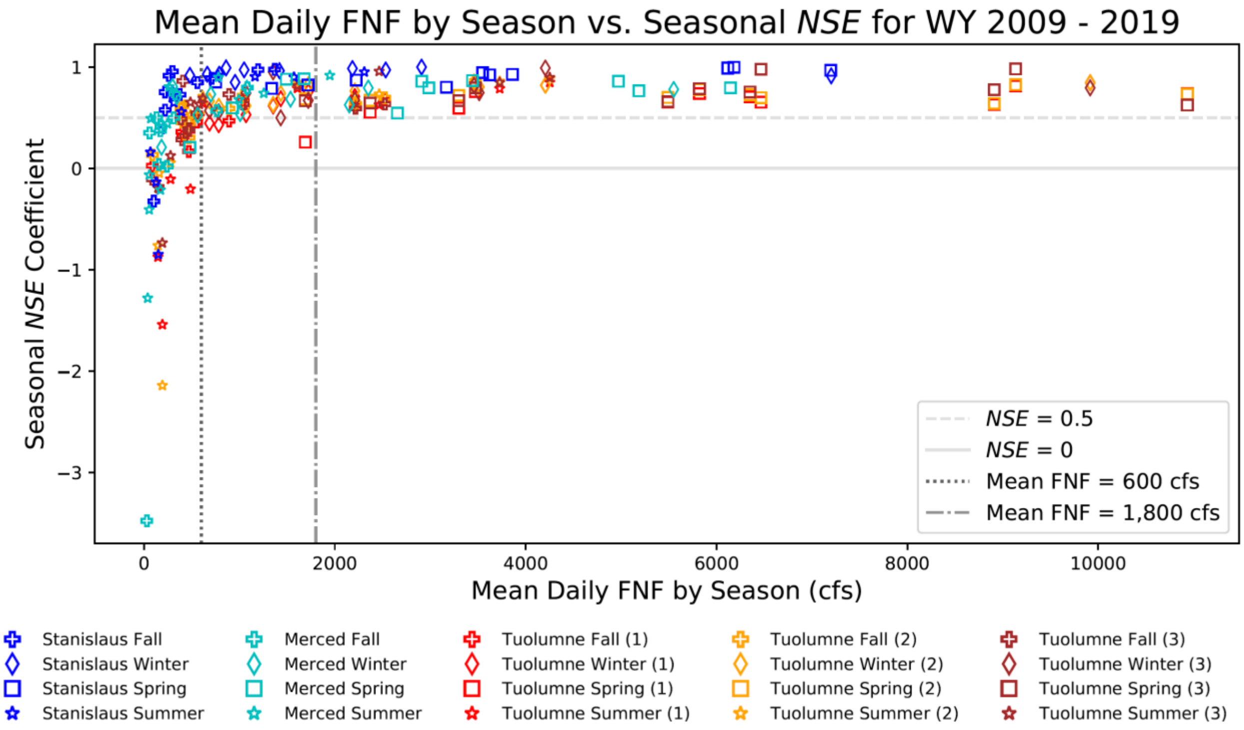


Figure 8: Mean Daily FNF by Season vs. Seasonal *NSE*, WY 2009-2019

This finding is further supported by results of the seasonal permutation test (seasonal *p*-Value). When the seasonal permutation test was applied, the Merced and Tuolumne Procedure 1 computations both failed to be statistically reproducible during the summer of WY 2014 (Figure 9). However, for all other computations the seasonal permutation test was inconclusive (Figure 9). The Stanislaus computation is not shown in Figure 9 because it was the only computation to receive a seasonal *p*-Value equal to zero for every season between WY 2009-2019. In Figure 9, the testing threshold of 0.05 is shown in grey.

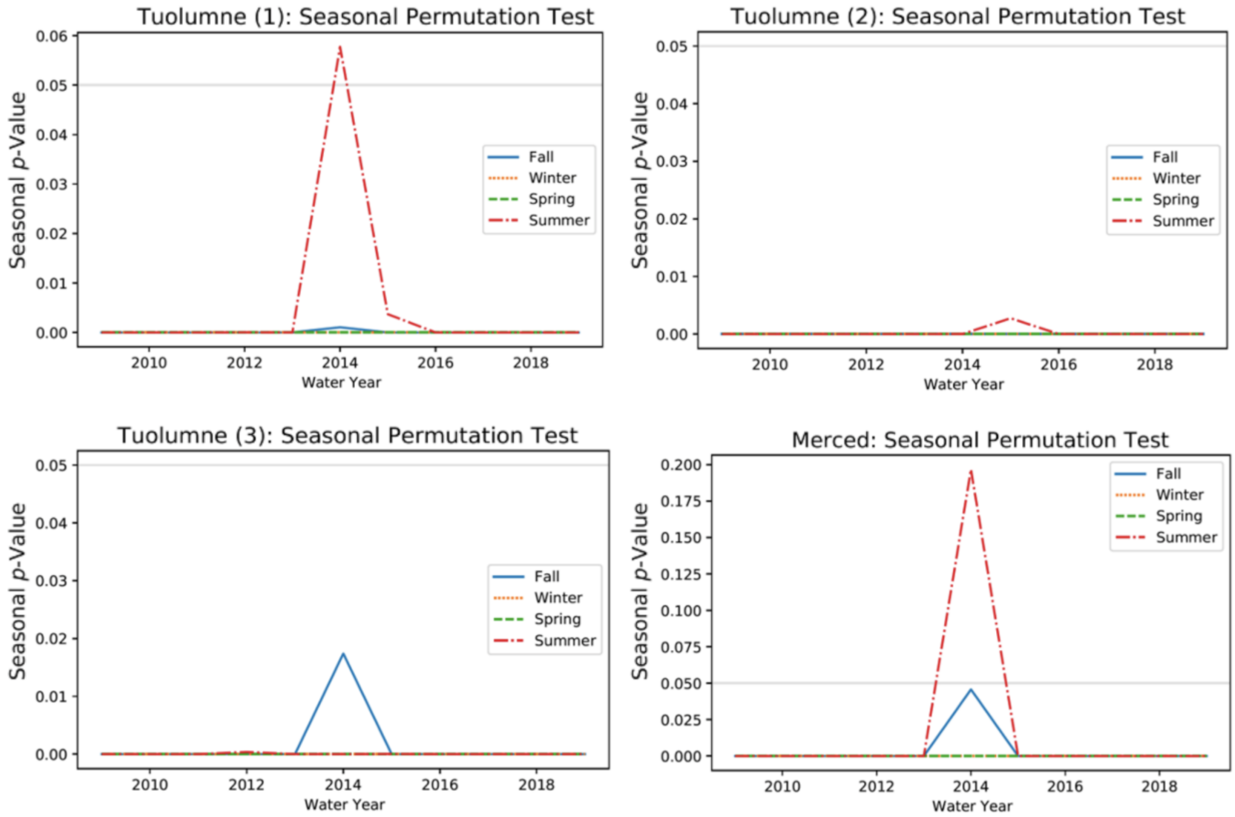


Figure 9: Seasonal *p*-Values by Water Year for Each FNF Calculation (Excluding the Stanislaus) for WY 2009-2019.

A secondary annual trend is shown by Figure 10: the average magnitude of daily discrepancies (*MAD*) computed by water year is highest during WY 2017, the wettest year of study. However, during WY 2017, the discrepancies were small compared to the mean FNF, and each computation received positive seasonal *NSE* and water year *NSE* scores (Figures 6 and 7). Additionally, during WY 2017 every computation received a zero seasonal and water year *p*-Value (Figure 9). This implies that all the reproducibility tests used in this analysis failed to identify any statistically non-reproducible daily FNF computations during WY 2017. From this, although the average magnitude of daily discrepancies was highest during WY 2017, these discrepancies did not significantly impact overall computation reproducibility.

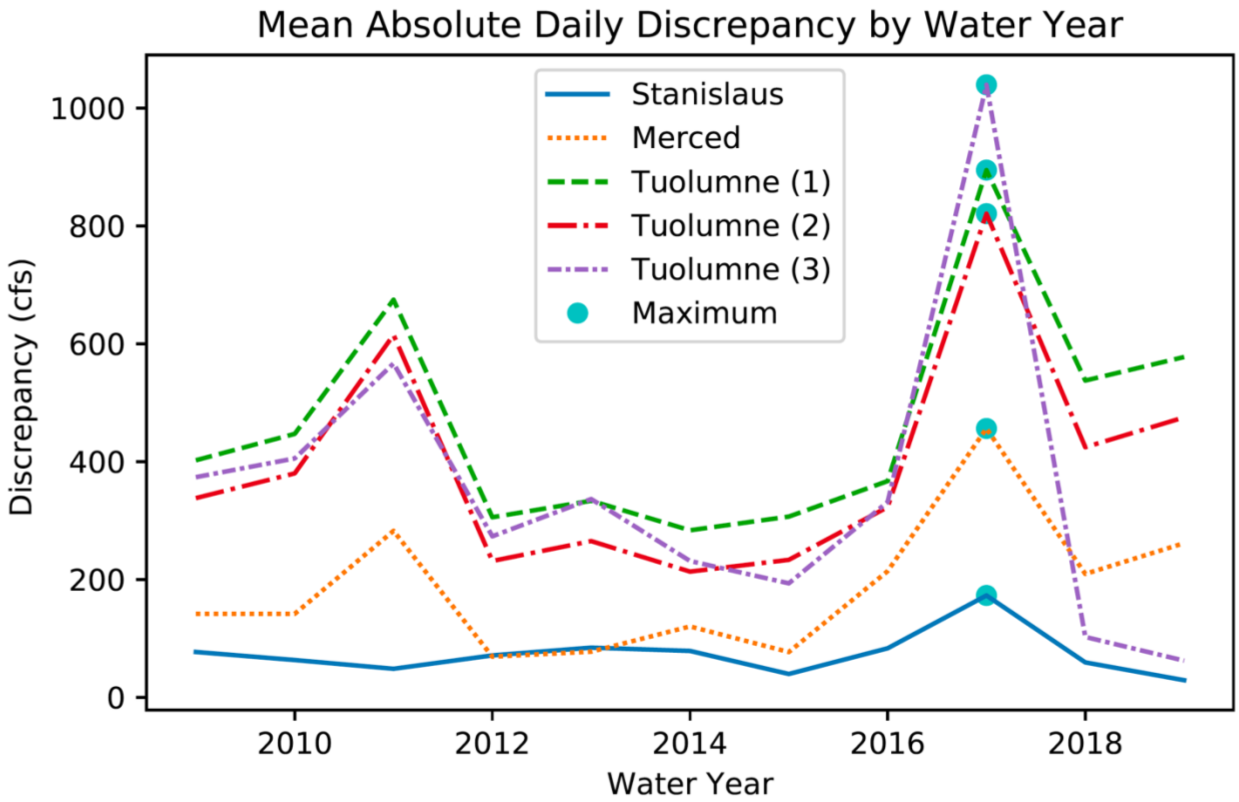


Figure 10: Mean Absolute Daily Discrepancy by Water Year for Each FNF calculation, WY 2009-2019

3.2. Stanislaus River

The combination of data sources that provided the best FNF reproduction for WY 2009-2019 used NWIS data for Tuolumne Canal and Lyons Reservoir storage, and CDEC data for all other impairments (Table 2). This combination of datasets provided the best reproduction of FNF out of all five computations of study, yielding an 11-year *NSE* value of 0.957 (Table 6).

Figure 11 shows a time series of reproduced and reported daily FNF values for the Stanislaus River from WY 2009-2019. Figure 12 gives a timeseries of daily discrepancies in the Stanislaus daily FNF computation.

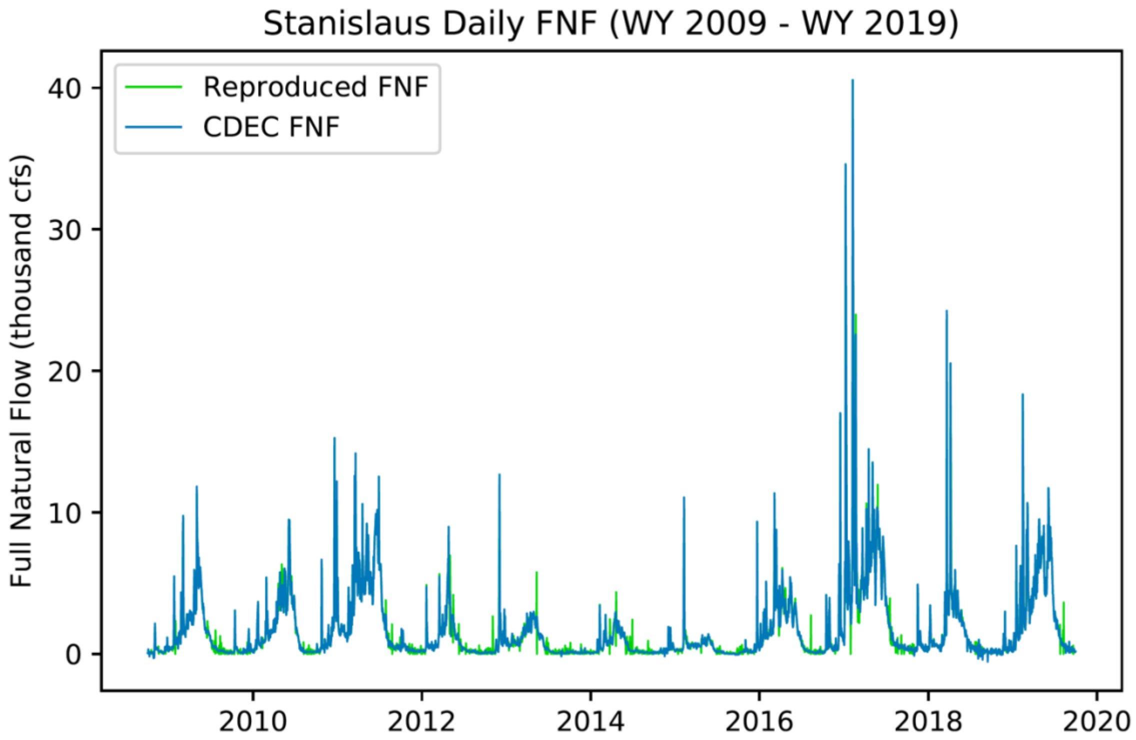


Figure 11: Reproduced and Reported Daily FNF for Stanislaus River, WY 2009-2019

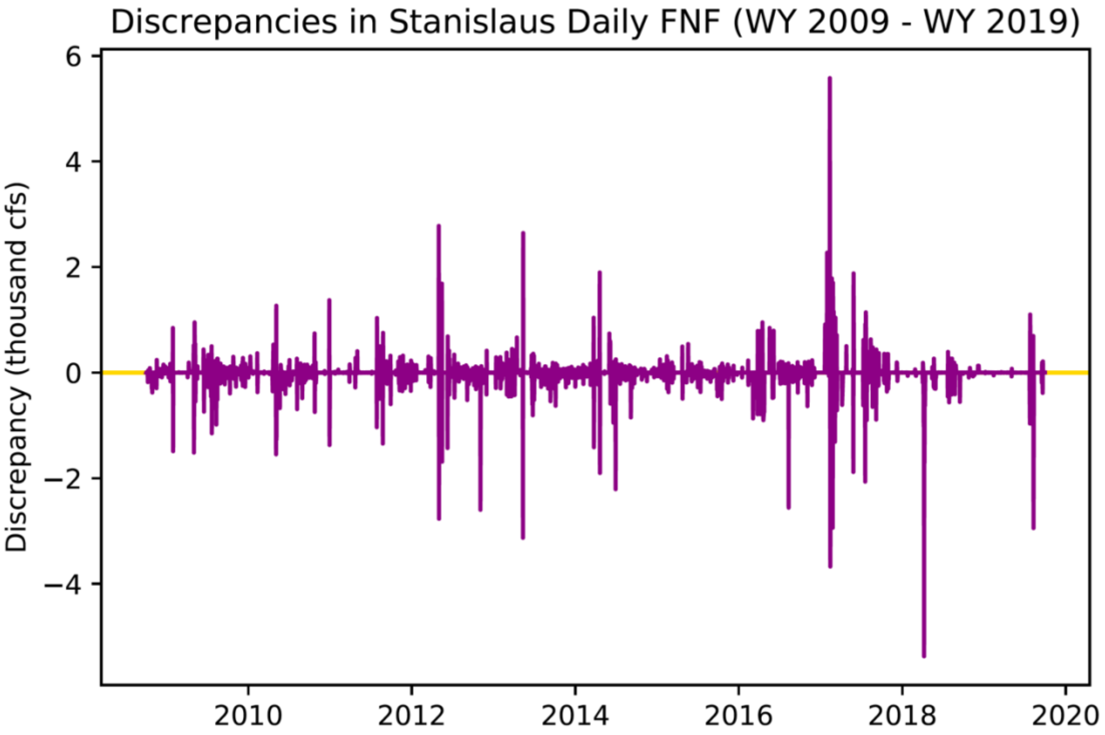


Figure 12: Daily Discrepancies in FNF Computation for Stanislaus, WY 2009-2019

3.3. Tuolumne River

All three computation methods (Procedure 1, Procedure 2, and Procedure 3) were used to reproduce the Tuolumne daily FNF computation. This section discusses the results of each computation independently, then finishes by comparing the performance of each method.

The combination of online data sources that yielded the best FNF reproduction for the Tuolumne Procedures 1 and 2 computations used NWIS data wherever possible, except for Don Pedro Reservoir and Lake Eleanor storage data where CDEC data provided the best fit (Table 3). This combination of datasets yielded an 11-year *NSE* value of 0.830 and 0.857 for Tuolumne Procedures 1 and 2 respectively (Table 6). It is surprising that CDEC data provided a better fit than NWIS for Lake Eleanor storage data because only WY 2019 is available on CDEC for this impairment. This is likely a result of the method used to choose input data sources; further development of this method might avoid choosing incomplete datasets.

Tuolumne Procedures 1 and 2

Figure 13 shows a time series of reproduced and reported daily FNF values for Tuolumne Procedure 1 for WY 2009-2019. Figure 14 gives a timeseries of daily discrepancies in the Tuolumne Procedure 1 daily FNF computation.

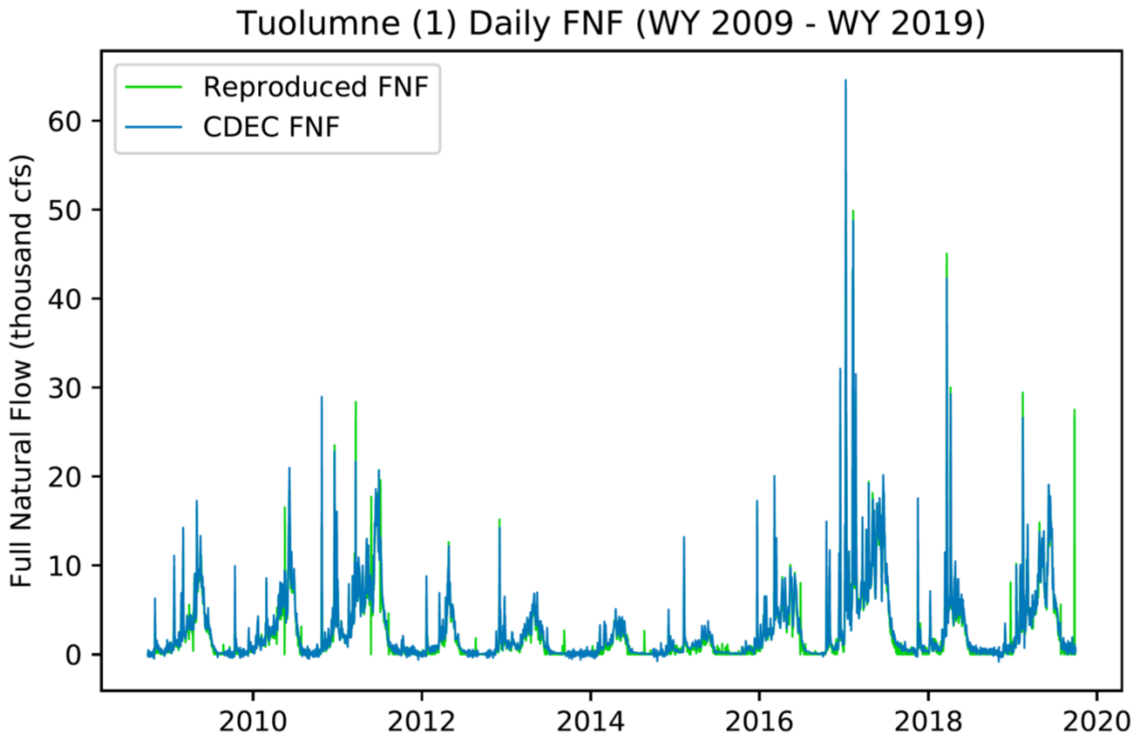


Figure 13: Reproduced and Reported Daily FNF for Tuolumne River Procedure 1, WY 2009-2019

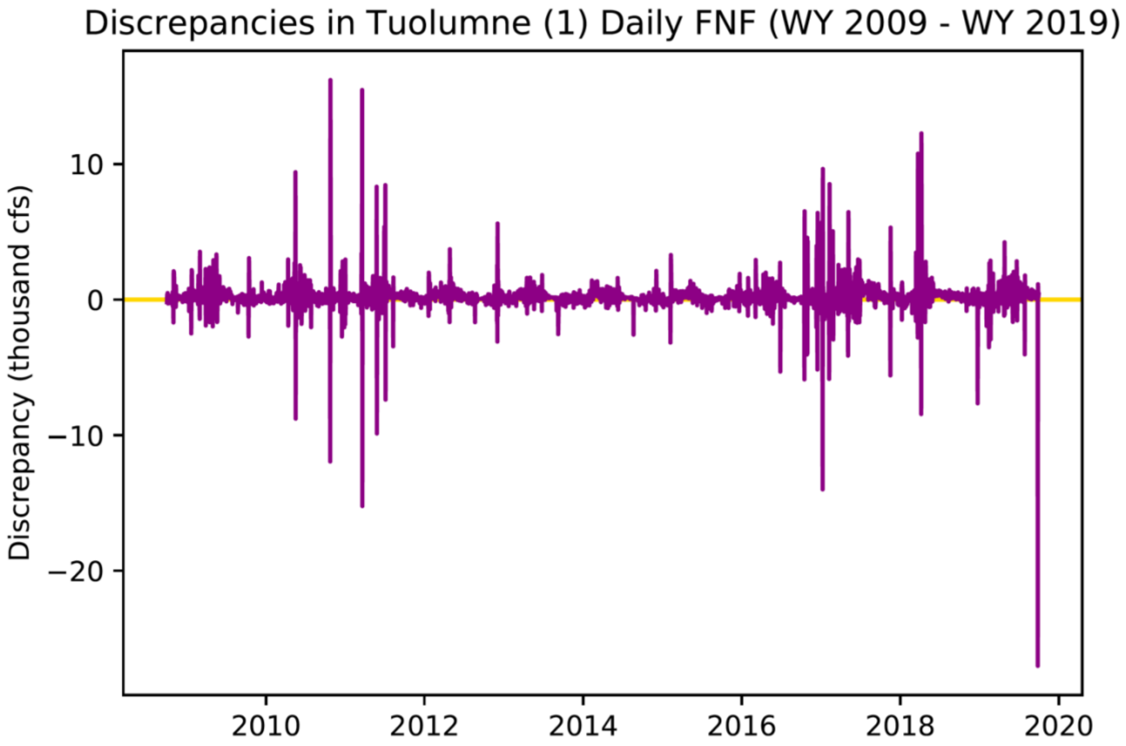


Figure 14: Daily Discrepancies in FNF Computation for Tuolumne River Procedure 1, WY 2009-2019

Note the high-magnitude discrepancy in September of 2019 (Figure 13; Figure 14). This discrepancy correlates with an outlier in the CDEC storage data for Don Pedro Reservoir.

The only difference between Tuolumne Procedures 1 and 2 were the inclusion of San Francisco pipeline diversion data in Procedure 2. Because of this, the timeseries of daily FNF and daily discrepancies for Procedures 1 and 2 are nearly identical, only Procedure 1 consistently underestimates daily FNF. Therefore, these timeseries plots are only provided for Procedure 1.

Tuolumne Procedure 3

Figure 15 shows a time series of reproduced and reported daily FNF values for Tuolumne Procedure 3 for WY 2009-2019. Figure 16 gives a timeseries of daily discrepancies in the Tuolumne Procedure 3 daily FNF computation.

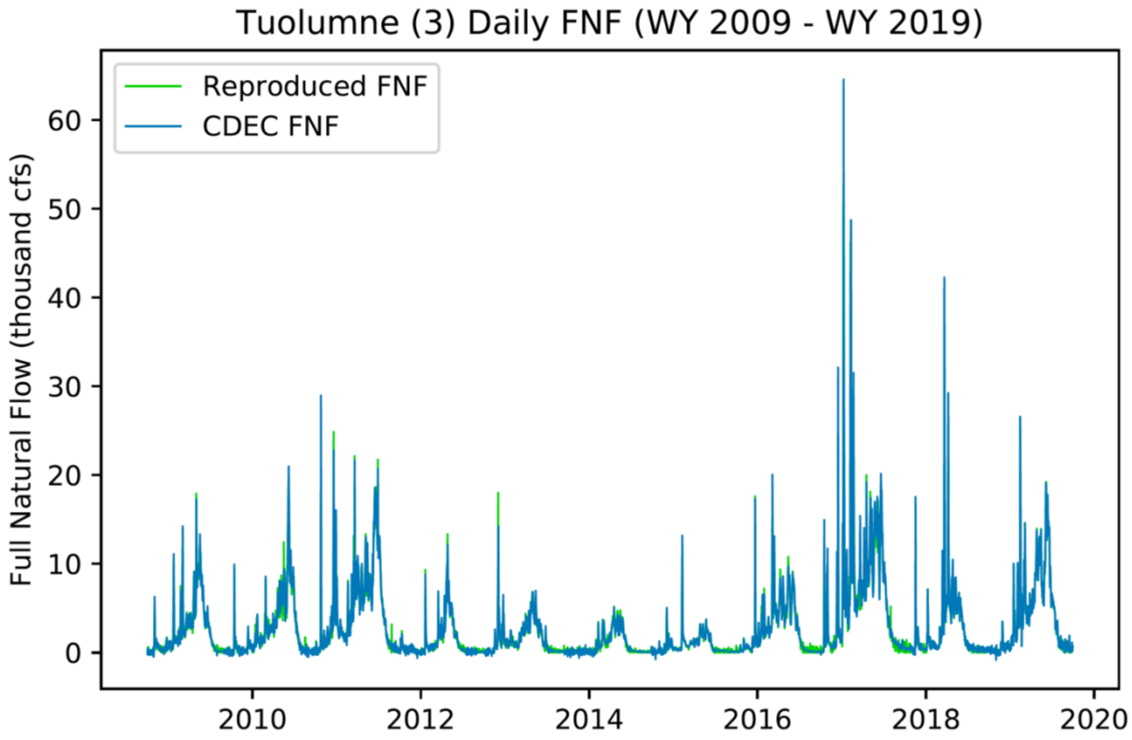


Figure 15: Reproduced and Reported Daily FNF for Tuolumne River Procedure 3, WY 2009-2019

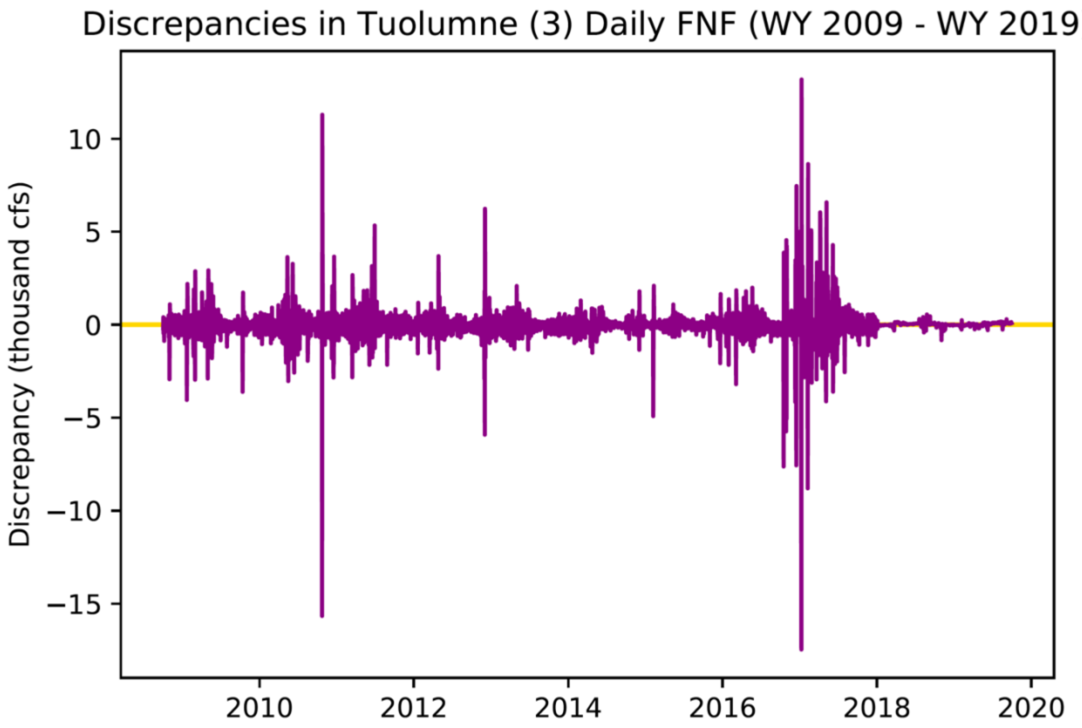


Figure 16: Daily Discrepancies in FNF Computation for Tuolumne River Procedure 3, WY 2009-2019

3.3.1. Tuolumne Method Comparison

Figure 17 shows that from WY 2009 to WY 2017, the Tuolumne Procedure 2 calculation outperforms Procedure 3 in seven of nine water years. This is likely because historical data supplied by Turlock ID for Procedure 3 has been systematically revised to better align with the NWIS (W. Monier, personal communication, January 16, 2020). In contrast, it is likely that most CDEC data used in Procedures 1 and 2 were not systematically revised (W. Monier, personal communication, January 16, 2020). This supports the theory that many discrepancies result from retrospective data correction.

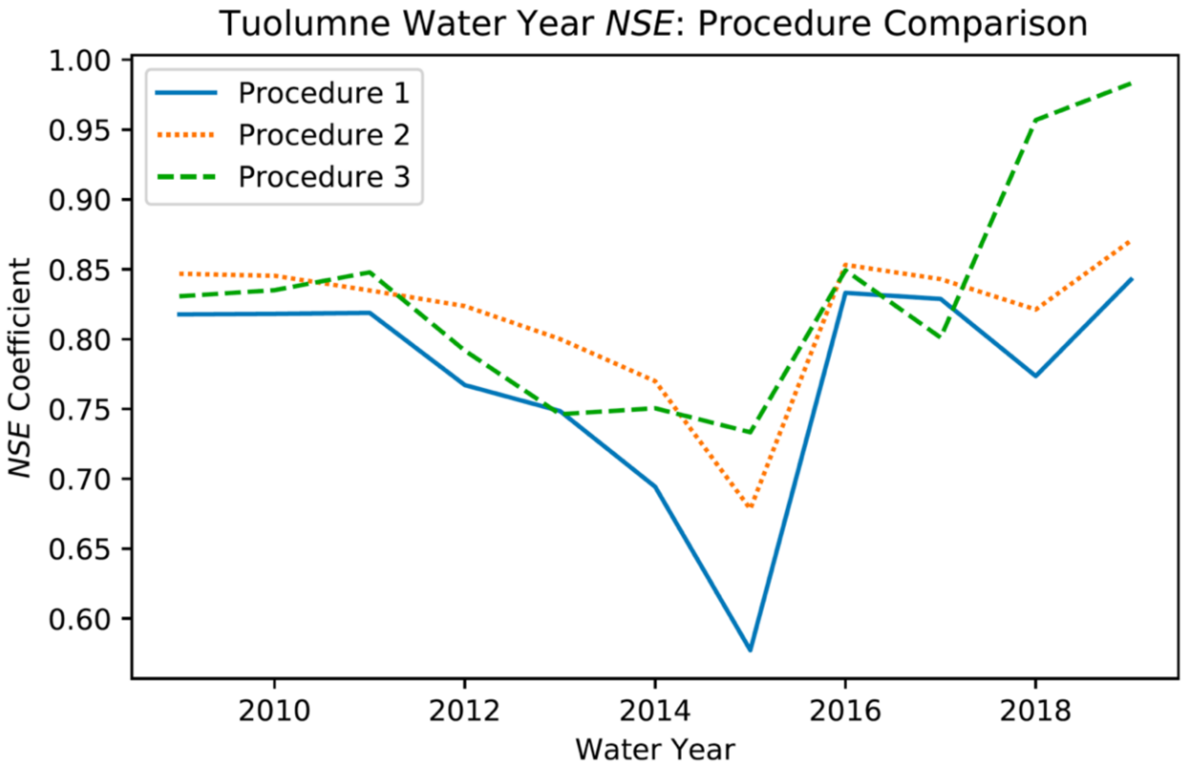


Figure 17: Water Year *NSE* for Each of the Tuolumne Daily FNF calculations: Procedures 1, 2, and 3

Additionally, when this analysis was conducted, daily FNF computing agency data for WY 2018-2019 supplied by Turlock ID were still under review (W. Monier, personal communication, January 16, 2020). This leads to the possibility that the two most recent years of data supplied by Turlock ID are provisional, causing Procedure 3 to outperform Procedure 2 for WY 2018 and WY 2019 (Figure 17). Differences between Procedures 2 and 3 show that access to the original (unrevised) source data improves FNF calculation reproducibility.

Figure 17 shows a second trend: Procedure 2 outperforms Procedure 1 every year. This is unsurprising since the only difference between Procedures 1 and 2 is inclusion of San Francisco pipeline diversion data in Procedure 2. Since Procedure 1 omits this element of the mass balance equation, Procedure 1 consistently underestimates FNF causing it to perform worse than Procedure 2 (Figure 17).

This consistent underestimation also accounts for trends in Figure 18: daily discrepancies for Procedure 1 are larger than for Procedure 2, in magnitude and value. For Procedure 1, the average magnitude of daily discrepancies (*MAD*) by water year are 44-113 cfs larger, and the average daily discrepancy by water year is between 197-274 cfs larger than for Procedure 2 (Figure 18). The differences between Procedures 1 and 2 show that data availability is a primary driver of FNF calculation reproducibility.

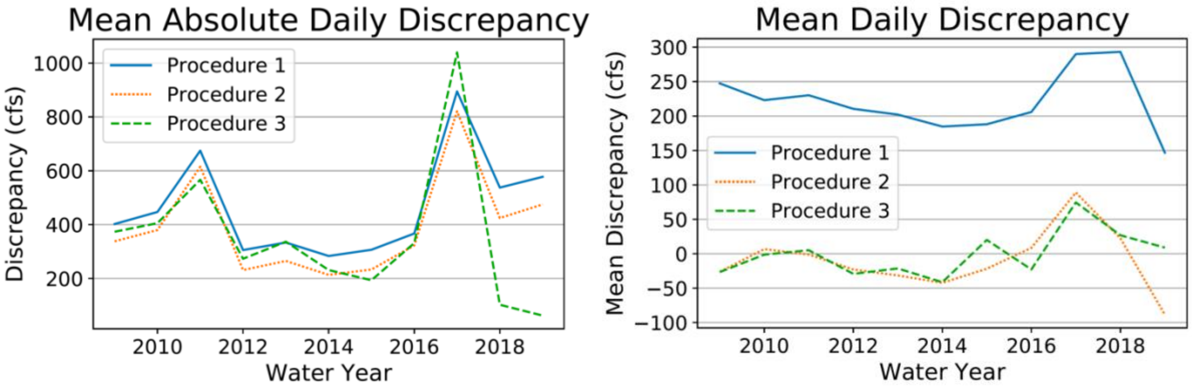


Figure 18: Mean Absolute Daily Discrepancy (left) and Mean Daily Discrepancy (right) by Water Year for the Tuolumne Daily FNF Calculation, Procedures 1, 2, and 3

3.4. Merced River

For the WY 2009-2019 Merced River daily FNF calculation, only Procedure 1 was reproduced because computing agency data were not available for the 11-year period of study. The combination of online data sources that provided the best FNF reproduction for the Merced daily FNF calculation used NWIS data for all three gages (Table 4). This combination of datasets yielded an 11-year *NSE* value of 0.870 (Table 6).

Figure 19 shows a time series of reproduced and reported daily FNF values for Merced River, WY 2009-2019. Figure 20 gives a timeseries of daily discrepancies in the Merced daily FNF computation.

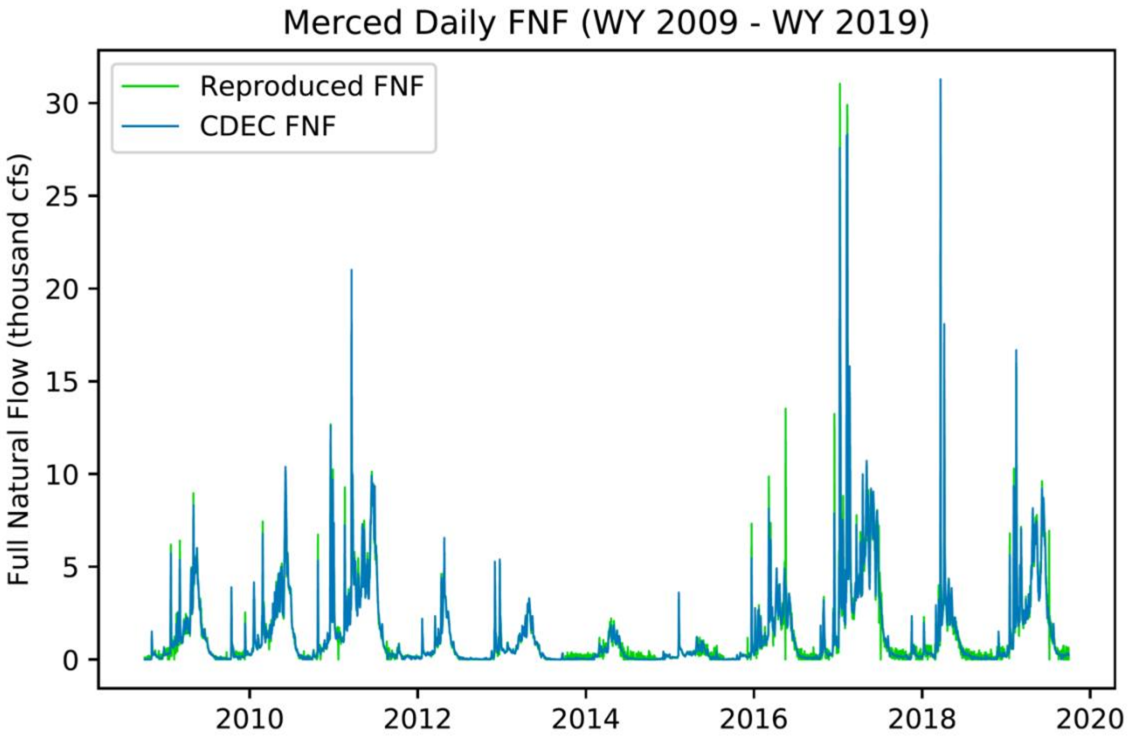


Figure 19: Reproduced and Reported Daily FNF for Merced River, WY 2009-2019

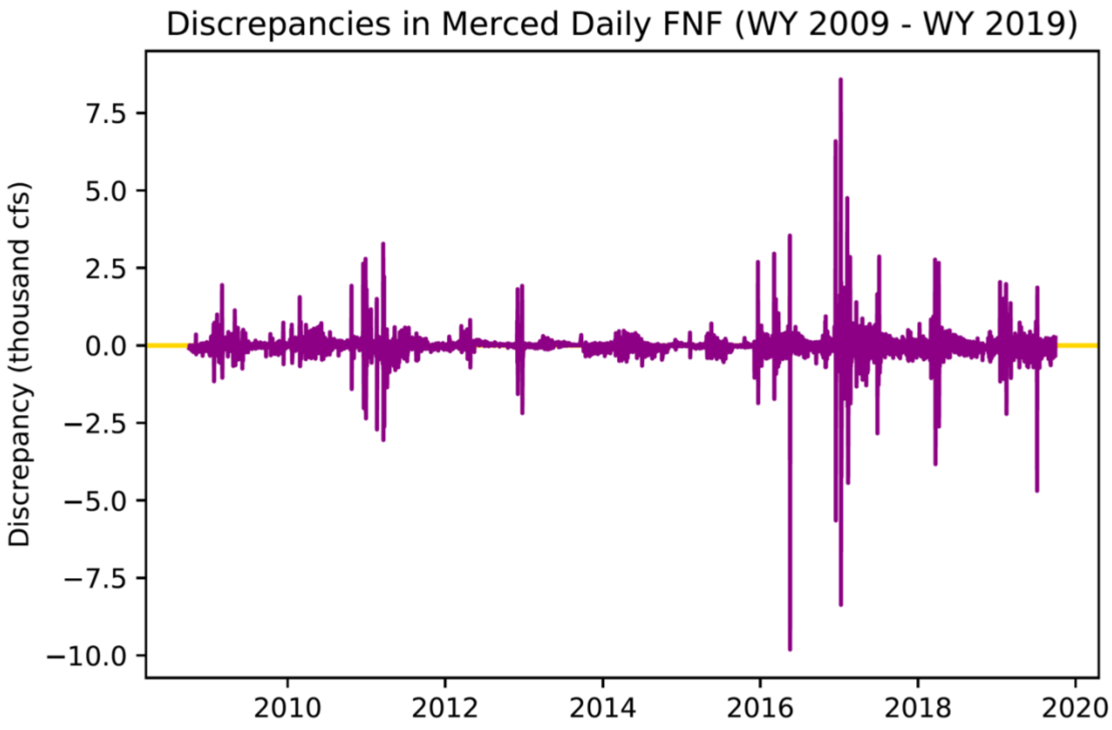


Figure 20: Daily Discrepancies in Daily FNF computation for Merced, WY 2009-2019

A notable characteristic of the Merced daily FNF calculation is that the summer and/or fall of WY 2014 received the lowest seasonal *NSE* values, and the highest seasonal *p*-Values and discrepancy significance values of any computation in this study. In other words, the dry seasons of the WY 2014 Merced daily FNF computation were the worst-performing seasons between WY 2009-2019 of all five computations in this study. For this reason, Figure 21 depicts reproduced and reported FNF values for the Merced daily FNF computation, WY 2014. In Figure 21, negative reproduced FNF values (shown in red) were replaced with zero before goodness-of-fit metrics were computed, as discussed in Section 2.4.2.

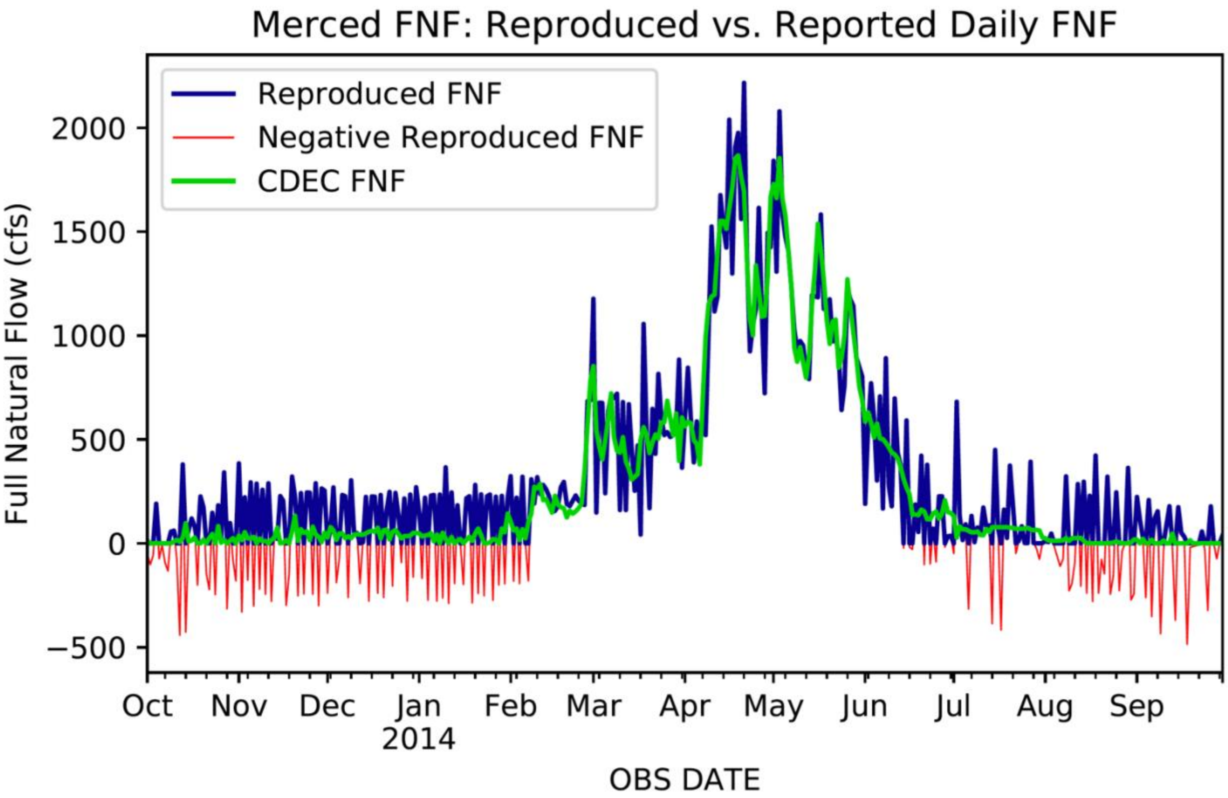


Figure 21: Reproduced and Reported Daily FNF for Merced River, WY 2014

From Figure 21we see that reproduced FNF values have more daily variability than FNF values posted to CDEC. Closer examination of one of the three components of the Merced daily FNF mass balance equation offers insight into the cause of this variability.

Figure 22 shows McClure Reservoir storage change data for WY 2014. From Figure 22, one can see that McClure Reservoir storage change data from NWIS oscillate between 1,000-acre foot intervals because these storage change data are rounded to the nearest thousand-acre foot. These oscillations correlate with compensating positive and negative discrepancies in the Merced daily FNF reproduction during WY 2014 (Figure 21). Therefore, it is likely that discrepancies in the WY 2014 reproduction of Merced daily FNF values are caused, at least in part, by the discretization of McClure Reservoir storage data posted to NWIS: input data used to reproduce daily FNF were rounded to the nearest thousand-acre foot, unlike data originally used to compute daily FNF. This further supports the theory that discrepancies between reproduced and reported daily FNF values are often caused by differences in institutional data management practices.

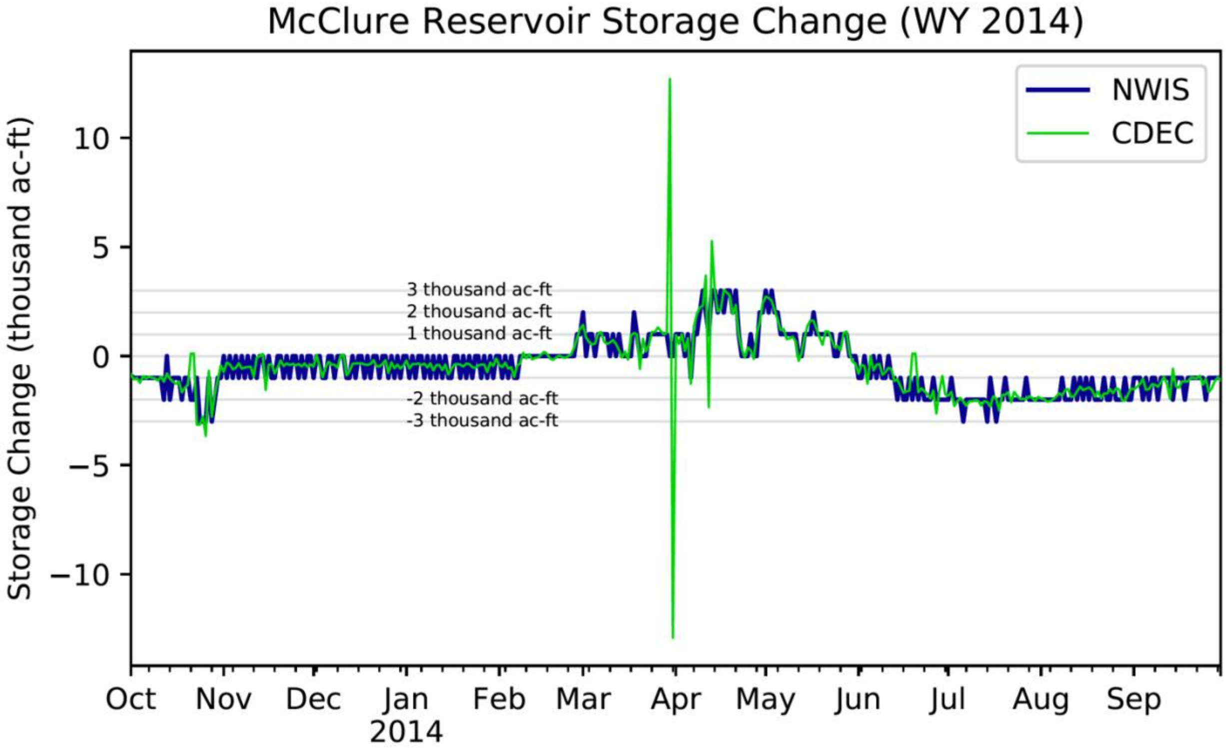


Figure 22: McClure Reservoir Storage Change from NWIS and CDEC, WY 2014

4. DISCUSSION

This section discusses potentially useful modifications to the analysis framework presented in this thesis. A baseline adjustment is introduced to show how the stringency of *NSE* can be increased without significantly changing observed trends in daily FNF reproducibility. Additionally, a rolling average is used to assess trends in daily FNF reproducibility when reproduced and reported daily FNF values are averaged on a rolling basis before they are compared. Finally, this section closes with a discussion on potential causes of discrepancies between reproduced and reported daily FNF values.

4.1. Baseline Adjustment

Recall from Section 2.7.1 (Metrics) that by the definition of the modified Nash-Sutcliffe coefficient of efficiency (Equation 4), the performance of each FNF reproduction is compared against the performance of the mean value of CDEC FNF computed over the entire time period of study. In this way, the mean value of CDEC FNF serves as a benchmark, or a baseline with which the model's performance is compared. For some applications of the *NSE*, the mean value of CDEC FNF computed over the entire record of study may not be a stringent enough baseline for the metric, particularly if there are strong seasonal trends in the dataset of study. To increase the metric's stringency, the *NSE* can be "baseline adjusted."

Equation 5 gives an example of the modified Nash-Sutcliffe Coefficient of Efficiency baseline adjusted by the four seasonal means of CDEC FNF computed for each water year. In Equation 5, *n* is the number of days in the period of study, O_i is the observed value on day *i* (CDEC FNF), P_i is the predicted value on day *i* (reproduced FNF), and $\underline{O}$' is the seasonal mean value of observed data associated with each day *i* in range 1 to *n*.

Equation 5: Baseline Adjusted Modified Nash-Sutcliffe Coefficient of Efficiency

$$\textit{Baseline Adjusted NSE} = 1 - \frac{\sum_{i=1}^{n} |O_i - P_i|}{\sum_{i=1}^{n} |O_i - \underline{O'}|}$$

For an example of how $\underline{O}$' would be computed, allow *i* to be any day in July of 2017. Then, the $\underline{O}$' associated with this day *i* would be equal to the mean value of daily CDEC FNF computed over the summer season of 2017. This adjustment establishes a new baseline time series for each water year: the four seasonal means of CDEC FNF. Table 7 provides the physical interpretation of the seasonally baseline adjusted *NSE* defined by Equation 5 when computed by water.

Table 7: Physical Interpretation of *NSE* Computed by Water Year, Baseline Adjusted by Season

	Physical Interpretation
Baseline Adjusted $NSE = 1$	For this water year, calculated FNF values perfectly reproduce daily FNF values reported on CDEC (*MAD* is equal to zero).
$0 <$ Baseline Adjusted $NSE < 1$	For this water year, calculated FNF values do not perfectly reproduce CDEC FNF, but calculated values perform better than four seasonal means of CDEC FNF.
Baseline Adjusted $NSE \leq 0$	For this water year, the four seasonal means of CDEC FNF better approximate daily FNF than the reproduced values.

To extend the example, we could further increase the stringency of the baseline adjusted *NSE* by redefining $\underline{O}$' in Equation 5 to be the *monthly* mean value of observed data associated with each day *i* in range 1 to *n*. Then, the new baseline time series for each water year would be the 12-monthly means of CDEC FNF. Table 8 shows that when computed over the entire record of study, each of these baseline adjustments make the coefficient more stringent than the 11-year *NSE*.

Table 8: 11-year *NSE* Baseline Adjustment Comparison (WY 2009-2019)

Daily FNF Computation	**11-year *NSE***	11-year *NSE* Baseline Adjusted by **Season**	11-year *NSE* Baseline Adjusted by **Month**
Stanislaus River	0.957	0.915	0.883
Merced River	0.870	0.743	0.637
Tuolumne River, Procedure 1	0.830	0.695	0.569
Tuolumne River, Procedure 2	0.857	0.743	0.637
Tuolumne River, Procedure 3	0.870	0.766	0.670

However, Figure 23 shows the same overall trends in reproducibility as Figure 6: reproducibility is lower during droughts years (WY 2013-2015), but not comparatively low during WY 2017. Therefore, we conclude that the additional stringency provided by monthly or seasonal baseline adjustments does not enhance the ability of the metric to identify overall trends in FNF reproducibility. Because of this, the

NSE coefficient defined by Equation 4 is sufficient to determine trends in overall daily FNF reproducibility.

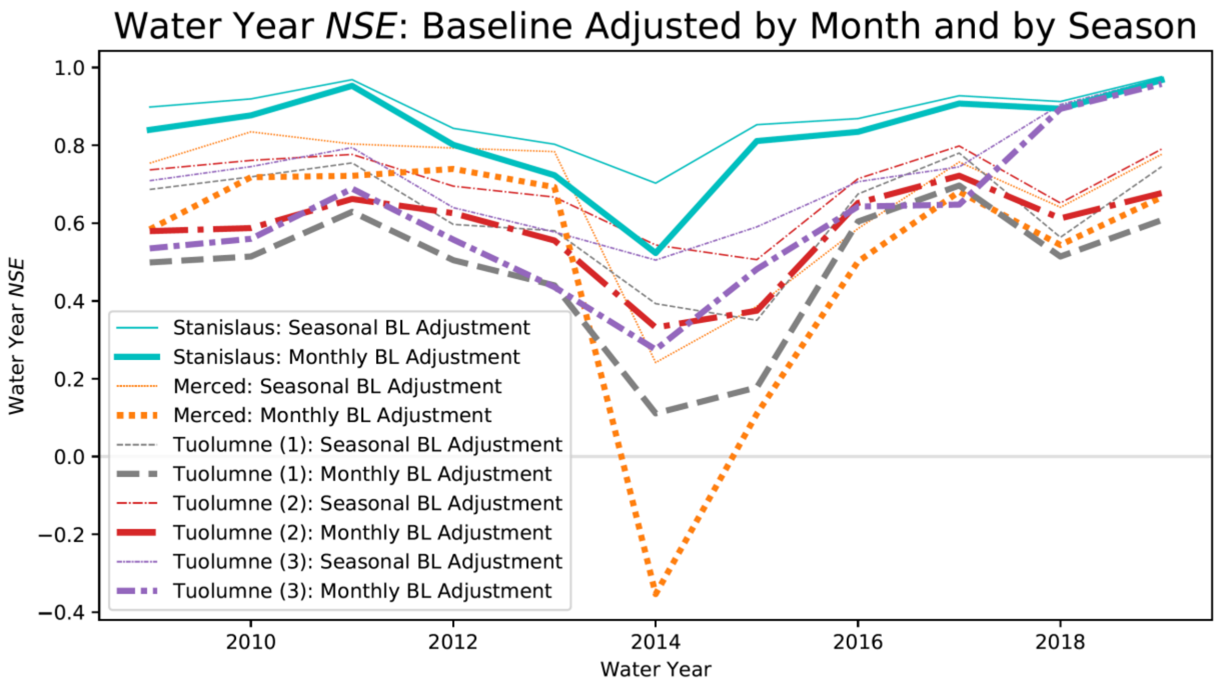


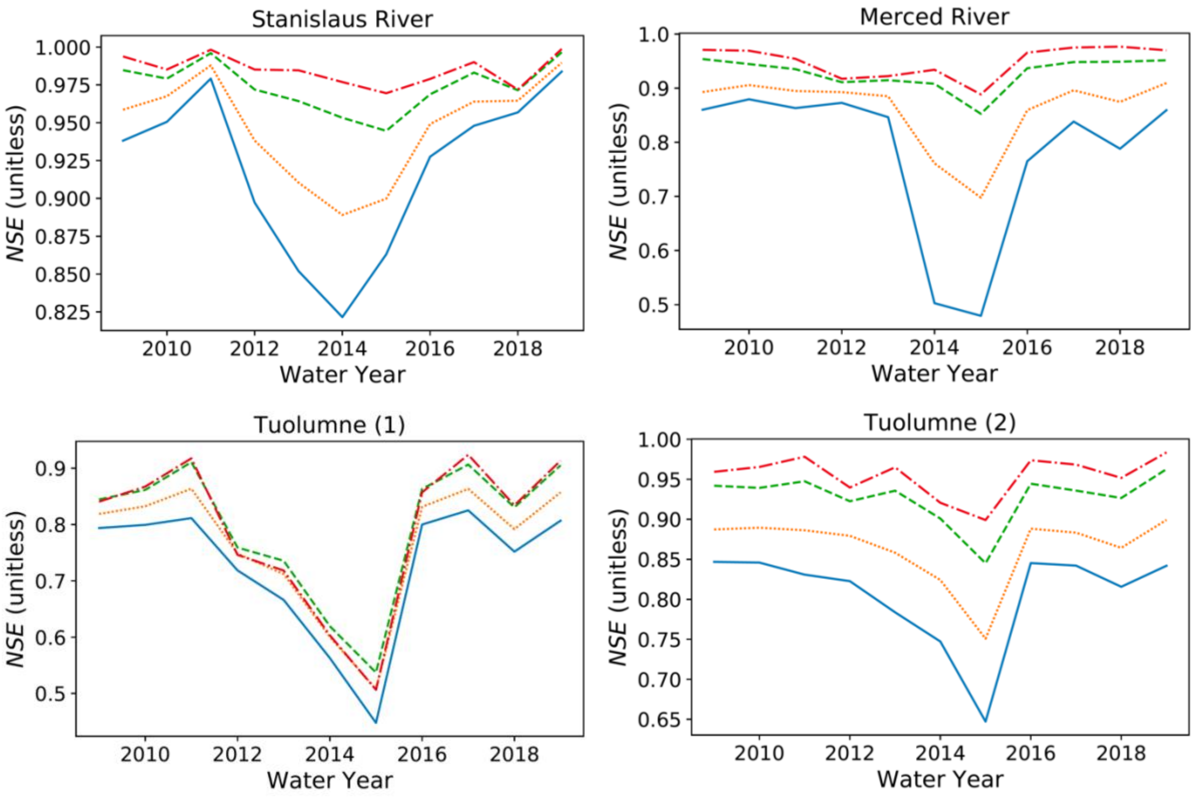
Figure 23: *NSE* Computed by Water Year, Baseline Adjusted by Season (Seasonal BL Adjustment) and by Month (Monthly BL Adjustment)

4.2. Rolling Average

From the daily discrepancy timeseries plots in the river-specific results (Sections 3.2 – 3.4), it appears that many large discrepancies occur in adjacent positive and negative pairs, perhaps due to sequential and compensating variations in storage change estimates. If daily FNF estimates are applied over time scales of two days or longer, these compensating variations might cancel out with multi-day averaging. To explore how multi-day averaging affects daily FNF reproducibility, reproduced and reported daily FNF timeseries were each averaged over two-, seven-, and 28-day rolling windows. In averaging the reproduced daily FNF timeseries, negative reproduced daily FNF values were not replaced with zero, as in the rest of this report. Instead, negative reproduced daily FNF values were allowed to compensate with adjacent over-estimations. Using the averaged reproduced and reported timeseries data, 11-year *NSE* and water year *NSE* values were computed for each FNF computation and rolling window. Results are presented in Table 9 and Figure 24.

Table 9: 11-year *NSE* Rolling Average Comparison (WY 2009-2019)

Daily FNF Computation	**11-year *NSE*:** no rolling average	**11-year *NSE*:** 2-day rolling average	**11-year *NSE*:** 7-day rolling average	**11-year *NSE*:** 28-day rolling average
Stanislaus River	0.957	0.968	0.984	0.991
Merced River	0.870	0.905	0.950	0.969
Tuolumne River, Procedure 1	0.830	0.839	0.872	0.877
Tuolumne River, Procedure 2	0.857	0.895	0.947	0.970
Tuolumne River, Procedure 3	0.870	0.909	0.958	0.979



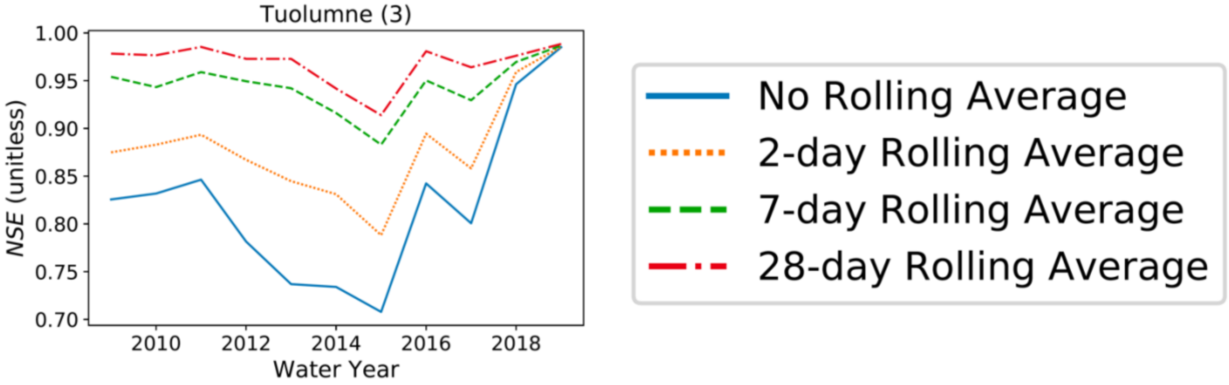


Figure 24: Water Year *NSE* Rolling Average Comparison (WY 2009-2019)

In general, taking the rolling average of both reproduced and reported daily FNF values improves the agreement between these two timeseries (Table 9, Figure 24). Even a two-day rolling average significantly increases water year and 11-year *NSE* values (Table 9, Figure 24). For every computation, 11-year *NSE* improves as the rolling average window is extended (Table 8). Similarly, longer rolling windows increased water year *NSE* values for most computations (Figure 24). However, the Tuolumne Procedure 1 computation preforms worse with a 28-day rolling average than the seven-day rolling average during drought years (WY 2012-2016). This is caused by the consistent underestimation of the Tuolumne Procedure 1 computation; longer rolling average windows decrease the mean absolute deviation of CDEC FNF more (proportionally) than the mean absolute discrepancy for the Procedure 1 computation during low flow time periods.

Though rolling averages improves overall reproducibility of daily FNF, rolling averages do not significantly change relative trends in reproducibility over time. The water year *NSE* values still decline during the WY 2012-2016 drought for each computation, regardless of the rolling average (Figure 24). And the Stanislaus computation is still the most reproducible computation of study with the highest 11-year *NSE*, while the Tuolumne River Procedure 1 computation was the least reproducible for each rolling window (Table 9).

4.3. Possible Cause of Discrepancies

In this Section, we use results presented in Section 3 to propose that discrepancies between reproduced and reported daily FNF values are caused by the use of currently available data instead of the provisional data originally used to compute CDEC FNF values.

First, consider trends in reproducibility discussed in Section 3.1: overall reproducibility of daily FNF values declines when flows are low (during drought years), even though the average magnitude of daily discrepancies was largest during the wettest water year of study (2017). The following tenet of hydrologic data collection and management is likely responsible for these trends: hydrologic monitoring equipment is least accurate when the volume of water measured is extremely high or extremely low. For example, geomorphological changes to a stream bed during a peak flow event could render rating tables for the stream reach erroneous, as rating table calibration depends on constant stream bed topography. This could motivate water managers to re-calibrate rating tables after peak flow events and retrospectively revise provisional gage flow data. This could explain the high-magnitude discrepancies observed in WY 2017.

Additionally, water managers are likely more economically motivated to ensure accurate data collection during drought periods, which could cause data revisions to be more frequent when flows are extremely low. This could account for the lack of reproducibility observed for each computation of study between WY 2013 – WY 2016.

Next, close examination of the daily FNF reproduction of Merced during WY 2014 revealed that differences in how input data are rounded is a potential cause of discrepancies (Section 3.4). In the case of the WY 2014 Merced computation, NWIS McClure Reservoir storage change data used in the reproduction of daily FNF was rounded to the nearest thousand ac-ft, while McClure Reservoir storage change data used in the original computation of CDEC FNF were likely not rounded to the nearest thousand ac-ft. This created compensating positive and negative discrepancies in the Merced daily FNF computation and negative reproduced daily FNF values during WY 2014 (Figure 21). This observation further supports the theory that discrepancies between reproduced and reported daily FNF values are caused by a lack of public access to the original source data used to compute daily FNF.

Finally, by comparing Procedures 2 and 3 in the Tuolumne River computation (Section 3.3), we showed that access to provisional, unrevised source data in WY 2018 and WY 2019 improved FNF calculation reproducibility for these two water years. From this, we concluded that discrepancies in the Procedures 2 and 3 Tuolumne daily FNF computations are caused, at least in part, by the adjustment of provisional data. It is likely that this finding can be generalized to other river basins since it is common practice to revise provisional hydrologic data to account for erroneous data points. However, further research is necessary to support this generalization.

Future work classifying the precise cause of specific discrepancies could illuminate possible physical implications of these discrepancies. For example, if a discrepancy is caused solely by the use of data rounded to the thousandth ac-ft, this discrepancy might not have a physical implication. This could be the case for Merced WY 2014. Alternatively, discrepancies caused solely by the retrospective correction of erroneous provisional data could have implications about the accuracy of CDEC FNF values, where time periods that are less reproducible could be less accurate. If one can prove that a specific set of discrepancies was caused solely by retrospective correction of inaccurate provisional data (*i.e.* if provisional data used in the original computation were collected by a poorly calibrated stream gage), this finding could inform the addition of a reconciliation step in the implementation of the SWRCB's instream flow requirement to adjust for erroneous water losses or gains incurred by regional water managers. Additionally, this finding could motivate the use of reproduced daily FNF values instead of CDEC FNF in the calibration of hydrologic simulation and forecasting models.

5. CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

The framework of analysis presented in this report was effective in identifying seasonal and multi-year trends in daily FNF reproducibility: for each computation of study, reproducibility declined during extremely low flow time periods, which are the most crucial time periods for instream flow requirements. This analysis also provided evidence to support the hypothesis that discrepancies between reproduced and reported daily FNF values are caused by a lack of public access to provisional data originally used to compute daily unimpaired flow values.

Additionally, the proposed reproducibility tests provide benchmarks with which to compare the reproducibility of a given FNF computation by season or water year. This might increase the usefulness of this framework for decisionmakers who seek to assess the potential risk of technical controversy or litigation from direct use of daily FNF values for contentious policy actions, like implementation of instream flow requirements in the SWRCB's Water Quality Control Plan (2018).

5.2. Future Work

As mentioned in Section 2.4.1, further research on computation-specific data management might improve reproducibility of daily FNF computations. Additionally, the method used to choose datasets for the daily FNF reproduction (Section 2.6) could be improved to avoid choosing incomplete datasets, as was the case for Lake Eleanor storage data in the Merced computation (Section 3.4).

One critique of this analysis framework that deserves further consideration is the possibility that the observed trends in reproducibility are a consequence of the metrics used. For example, low flows are likely to correlate with small mean absolute deviation values, which could artificially lower the *NSE* coefficient during droughts. Additionally, by design, low flows inflate the ratio of *MAD* to mean CDEC FNF. And if a river went dry for a significant portion of a dry season, this could result in artificially inflated seasonal *p*-Values as permuting an array of mostly zeros might not significantly change the timeseries.

Finally, further research on the precise cause of specific discrepancies might illuminate the physical implications of daily FNF reproducibility on both the SWRCB's instream flow requirements and on hydrologic management and modeling in general. This future research might make use of CDEC's revised data point flags – a dataset that was not considered in this analysis.

REFERENCES

Bardsley, W. E., & Purdie, J. M. (2007). An invalidation test for predictive models. Journal of hydrology, 338(1-2), 57-62. https://www.sciencedirect.com/science/article/abs/pii/S0022169407001151

California Department of Water Resources, Bay-Delta Office (DWR-BDO). (2016). Estimates of Natural and Unimpaired Flows for the Central Valley of California: Water Years 1922-2014, Draft. Sacramento, CA: California Department of Water Resources, Bay-Delta Office. https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/california_waterfix/exhibits/docs/petitioners_exhibit/dwr/part2_rebuttal/dwr_1384.pdf

California Department of Water Resources, Division of Flood Management (DWR-DFM). (2016 Revised). Derivation of Unimpaired Runoff in the Cooperative Snow Surveys Program. Sacramento, CA: California Department of Water Resources, Division of Flood Management.

Legates, D. R., and McCabe, G. J. (1999), Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35(1), 233– 241, doi:10.1029/1998WR900018. https://www.researchgate.net/publication/235810981_Evaluating_the_Use_Of_Goodness-of-Fit_Measures_in_Hydrologic_and_Hydroclimatic_Model_Validation

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*, 885. https://swat.tamu.edu/media/1312/moriasimodeleval.pdf

Pulido, A., Pasner, Y., Murdoch, L., Lund, J., & Fleenor, W. (2021). Unimpaired Flow Calculations for Three San Joaquin Basin Tributaries in California's Central Valley (Draft). University of California, Davis.

State Water Resources Control Board. (2006). Water Quality Control Plan for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary. Sacramento, CA: State Water Resources Control Board, Division of Water Rights. https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/wq_control_plans/2006wqcp/

State Water Resources Control Board. (2018). Water Quality Control Plan for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary. Sacramento, CA: State Water Resources Control Board, Division of Water Rights. https://www.waterboards.ca.gov/plans_policies/docs/2018wqcp.pdf