

UC Davis

San Francisco Estuary and Watershed Science

Title

A Survey of X2 Isohaline Empirical Models for the San Francisco Estuary

Permalink

<https://escholarship.org/uc/item/8x7358kq>

Journal

San Francisco Estuary and Watershed Science, 19(4)

Authors

Rath, John S.
Hutton, Paul H.
Ateljevich, Eli S.
et al.

Publication Date

2021

DOI

10.15447/sfew.s.2021v19iss4art3

Supplemental Material

<https://escholarship.org/uc/item/8x7358kq#supplemental>

Copyright Information

Copyright 2021 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

RESEARCH

A Survey of X2 Isohaline Empirical Models for the San Francisco Estuary

John S. Rath¹, Paul H. Hutton¹, Eli S. Ateljevich², Sujoy B. Roy^{*1}

ABSTRACT

This work surveys the performance of several empirical models, all recalibrated to a common data set, that were developed over the past 25 years to relate freshwater flow and salinity in the San Francisco Estuary (estuary). The estuary's salinity regime—broadly regulated to meet urban, agricultural, and ecosystem beneficial uses—is managed in spring and certain fall months to meet ecosystem objectives by controlling the 2 parts per thousand bottom salinity isohaline position (referred to as X2). We tested five empirical models for accuracy, mean, and transient behavior. We included a sixth model, employing a machine learning framework and variables other than outflow, in this survey to compare fitting skill, but did not subject it to the full suite of tests applied to the other five empirical models. Model performance was observed to vary with hydrology, year, and season, and in some cases exhibited unique limitations as a result of mathematical formulation. However,

no single model formulation was found to be consistently superior across a wide range of tests and applications. One test revealed that the models performed equally well when recalibrated to a uniformly perturbed input time-series. Thus, while the models may be used to identify anomalies or seasonal biases (the latter being the subject of a companion paper), their use as inverse models to infer freshwater outflow to the estuary from salinity observations is not expected to improve upon the absolute accuracy of existing outflow estimates. This survey suggests that, for analyses that span a long hydrologic record, an ensemble approach—rather than the use of any individual model on its own—may be preferable to exploit the strengths of individual models.

KEY WORDS

San Francisco Estuary, Sacramento–San Joaquin Delta, estuarine salinity, X2, empirical model

INTRODUCTION

Salinity intrusion in estuarine and deltaic waters is a natural phenomenon, although in developed watersheds the extent and timing can be heavily influenced by freshwater withdrawals and upstream water management and use. Excessive salinity intrusion is widely reported, notably in drier climates (Alber 2002). Salinity intrusion can

SFEWS Volume 19 | Issue 4 | Article 3

<https://doi.org/10.15447/sfews.2021v19iss4art3>

* Corresponding author: sujoy.roy@tetrattech.com

1 Tetra Tech, Inc.
Lafayette, CA 94549 USA

2 California Department of Water Resources
Sacramento, CA 95814 USA

adversely affect native estuarine ecosystems as well as human uses of freshwater in proximity to estuaries and has therefore been subject to study and management in many parts of the world (Sklar and Browder 1998; Murray Darling Basin Ministerial Council 1999; Reinert and Peterson 2008; Fernández-Delgado et al. 2007). This is particularly true in California, with its relatively dry and variable (both intra- and inter-annually) Mediterranean climate. San Francisco Estuary (the estuary), composed of a series of interconnected embayments, rivers, sloughs, and marshes as well as the Sacramento-San Joaquin Delta (referred to hereafter as the Delta), is an ecosystem of vital international importance, and is also the location of several diversion facilities that export freshwater to agricultural and municipal users across large parts of the state (Kimmerer 2004; Luoma et al. 2015). Water management has been concerned with salinity intrusion adversely affecting water supply and agriculture in the Delta since the early part of the 20th century (CDPW 1931), and in more recent decades these concerns have evolved to include effects on the native ecosystem. Over the past 8 decades, beginning with the construction of major upstream surface reservoirs, the salinity regime of the Delta has been highly managed. During most of the year, reservoir releases and freshwater exports are controlled to maintain downstream salinity targets based on water supply and ecosystem goals (Lund et al. 2010; Luoma et al. 2015).

Salinity intrusion into the Delta varies as a function of freshwater outflow and tidal mixing, ranging from near-ocean salinity at the mouth of San Francisco Bay at Golden Gate, to near-freshwater salinity in the upstream Delta channels (Figure 1), with an intermediate zone of freshwater influence where salinity moves landward and seaward seasonally as a function of Delta outflow. Other drivers of mixing include wind forcing, barometric pressure, and coastal effects (such as sea level and currents); however, these drivers are not typically considered in empirical models that we present here. The low salinity zone occurs at the landward edge of the estuary, where average daily salinities range from

approximately 1 to 6 parts per thousand (ppt). X2—defined as the daily average position of the 2 ppt bottom salinity isohaline, and measured as the distance in kilometers from Golden Gate—is a common indicator of the location of the low salinity zone, and has been correlated with the abundance of several estuarine species (Jassby et al. 1995). The position of the X2 isohaline during the months of February through June has been used as a basis of flow management in the Delta since 1995 and is regulated under the California State Water Resources Control Board's Water Right Decision D-1641 (CSWRCB 2000). Currently, the X2 position is also being managed during September and October at the end of wet, above-normal, and below-normal water years (WYs). Water years in California begin on October 1 of the previous calendar year. A large and growing published literature on the use of X2 position as a variable in fish abundance relationships will continue to shape future management of the estuary (e.g., Jassby et al. 1995; Feyrer et al. 2007; Kimmerer et al. 2009, 2013; Mac Nally et al. 2010; Cloern et al. 2017; Tamburello et al. 2019; Murphy and Weiland 2019).

Aside from its regulatory role, X2 also functions as a compact metric for describing the salinity distribution in the estuary. Jassby et al. (1995) observed that X2 “collapses” salinity data about an equilibrium mean salinity distribution and concluded that one can infer the entire mean salinity distribution if the isohaline position is known. Monismith et al. (2002) elaborated on the approximate self-similar characteristics of the estuary's mean salinity distribution, and given assumptions of near self-similarity, Hutton et al. (2015) proposed an empirical model that predicts salinity as a function of X2 and longitudinal distance along the estuary.

Given the key role of X2 position in Delta water management and to California water resources in general, considerable effort has been devoted to understanding X2 behavior under various conditions, notably as a function of freshwater outflow and coastal water level. The seminal work that defined X2 and proposed its use as a regulatory metric, published in Schubel (1993),

included a detailed description of an empirical model that related observed salinity and Delta outflow. This model, subsequently published in Jassby et al. (1995), is well established in practice and has been shown to possess a balance of general accuracy, parsimony, and ease of implementation.

Over the following 25 years, several other empirical X2 models were published. Motivations for these efforts varied, and included an aspiration for greater empirical accuracy, a need to evaluate historically low flow conditions, and a desire to explore particular hypotheses concerning estuarine circulation and outflow. Examples of the latter motivation include the appropriate power law relationships to apply over the estuary's diverse terrain, and whether the time rate of the estuarine response is more proportional to flow or salinity (which has implications for system response at fortnightly frequencies). The empirical models assume relationships between freshwater flow and intrusion length that are formally simple, ignoring system complexities such as estuarine geometry and the tidal operation of the nearby Suisun Marsh Salinity Control Structure (see [Figure 1](#)) that reroutes flow in the vicinity of the Sacramento and San Joaquin rivers' confluence. The models were generally calibrated with observed salinity monitoring data in the estuary (Monismith et al. 2002; Hutton et al. 2015, Monismith 2017; Rath et al. 2017); however, one model (MacWilliams et al. 2015) was calibrated with salinity data from a spatially detailed hydrodynamic model. Because these models were published over a lengthy time horizon, different data sets and time-periods were used in their development. Although the Jassby et al. (1995) model arguably remains the most widely used empirical X2 model, to our knowledge, no comprehensive comparison of the published models has been undertaken.

The research objectives of this work include evaluating and comparing the performance of the different empirical X2 model formulations under a wide range of hydrologic conditions; testing their steady state, mean, and transient responses

at different flow rates; and evaluating whether such models can be effectively used inversely as tools for estimating freshwater outflow from a known X2 position. This last objective is of interest because the measurement of net freshwater outflows from the Delta is challenged by large tidal flows in the region of the river confluence (especially under low flow conditions). Others have hypothesized that Delta outflow can be effectively estimated from the salinity state or gradient, which can be more easily measured in the field (CDWR 2016; Fleenor et al. 2016).

To meet these research objectives, we first recalibrated the empirical models using outflow and X2 data from a common and relatively recent 10-year period of record. Recalibration was a key part of this study and involved constraining all models with the same flow and X2 data to ensure that differences among models described in this work were the result of the model formulation and not the period of data used for calibration, which in many cases were decades apart from one another. We recalibrated the models to place them all on the same footing, and to ensure that differences in behavior could be attributed to the formulation and not the particular data set embedded in the original published calibrations. Therefore, it is important to note that when we use the term "model" in this analysis, we are not referring to the equations and parameters as published, but rather to recalibrated versions of the original empirical formulations.

We examined recalibrated model performance by comparing predictions with X2 position (using a variety of statistical methods) as estimated from a longer 50-year observed salinity record. Next, we examined recalibrated model residual trends over the longest available observed salinity record (that spanned nearly a century) to determine if the estuary's flow-salinity response appears significantly altered over the past century. We also explored steady state and dynamic model responses to step changes and oscillations in outflow, using tests analogous to ones applied by Monismith (2017). Finally, we subjected the recalibrated models to outflow time-series perturbations, to evaluate whether they can be

effectively used inversely as tools for estimating freshwater outflow from a known X2 position. This survey of published empirical X2 models for the estuary provides insight into their behavior under typical and extreme conditions relevant for management and planning activities, and also provides insight into their broader utility in understanding the estuary's flow-salinity response.

BACKGROUND

Geographic and Physical Setting

The geographic focus of this paper is the upper portion of the estuary, including Carquinez Strait, Suisun Bay, and the western edge of the Delta (Figure 1). The Delta is the entry point of over 90% of the freshwater inflow to San Francisco Bay (Cheng et al. 1993) with inflow primarily from the Sacramento and San Joaquin rivers. The basic conceptual model of freshwater-saltwater mixing in estuaries with seasonally varying flow patterns such as the estuary is as follows: freshwater flows repel salinity downstream (seaward) across a mixing zone (with longitudinal and vertical gradients), and saltwater intrudes upstream (landward) during periods of low freshwater flow. The extent of the salinity gradient varies with tides on hourly to daily time-scales, and varies with the time history of freshwater flows on daily to seasonal time-scales. Salinity management in the estuary is primarily concerned with daily, fortnightly, and seasonal variability.

Definition of X2

X2, a measure of intrusion length, was originally defined in terms of bottom salinity, i.e. the position of the 2 ppt bottom salinity isohaline, and measured as the distance in kilometers from Golden Gate along the estuary centerline (Schubel 1993). However, the body of work on model development, model application, and regulatory compliance generally relies on surface salinity measurements computed from specific conductance. Use of surface salinity as a surrogate for bottom salinity is facilitated by an abundance of surface salinity measurements throughout the estuary, which developed in part because of precedent and the operational

challenges of maintaining salinity sensors at depth. The estuary is known to be vertically stratified, with increasing stratification at greater river flows (Monismith et al. 2002). Jassby et al. (1995) accommodated stratification by using a constant factor to relate the bottom salinity to surface salinity, i.e. 2 ppt bottom salinity is assumed to correspond to 1.76 ppt surface salinity. MacWilliams et al. (2015) discussed this assumption and evaluated its accuracy and limitations; further examination of this assumption was beyond the scope of our work. We note that in some previous work, empirical model calibrations were performed using observed bottom salinity measurements (of which a more limited data set exists) or synthetic data from spatially resolved numerical models. When X2 is estimated from observed data at fixed locations, the value is computed by interpolation; however, a standardized interpolation methodology has not been identified within the scientific community, and the stations vary with respect to their representativeness laterally across the estuary. Even the definition of axial distance in various published works lacks a prescribed set of landmarks or routes. Based on variations in the authors' own work and that of the referenced models, this introduces approximately 2 km of uncertainty to intrusion length estimates. For example, details on the route along the San Joaquin River upstream of the confluence—through New York Slough or Broad Slough (Figure 1)—are occasionally omitted. When X2 is estimated from a numerical model, the value is obtained directly from a spatially resolved grid. For regulatory purposes, X2 is defined as the position of the 2.64 milliSiemens per cm (mS cm^{-1}) surface isohaline (CSWRCB 2000).

Published models and analyses often define the X2 isohaline, when located upstream (i.e., east) of the Sacramento and San Joaquin rivers' confluence, to reflect an average position along the river branches. Hutton et al. (2015), by reporting unique model fits for each river branch, demonstrated greater salinity intrusion along the San Joaquin River branch for a given outflow. For consistency with other published work, here we

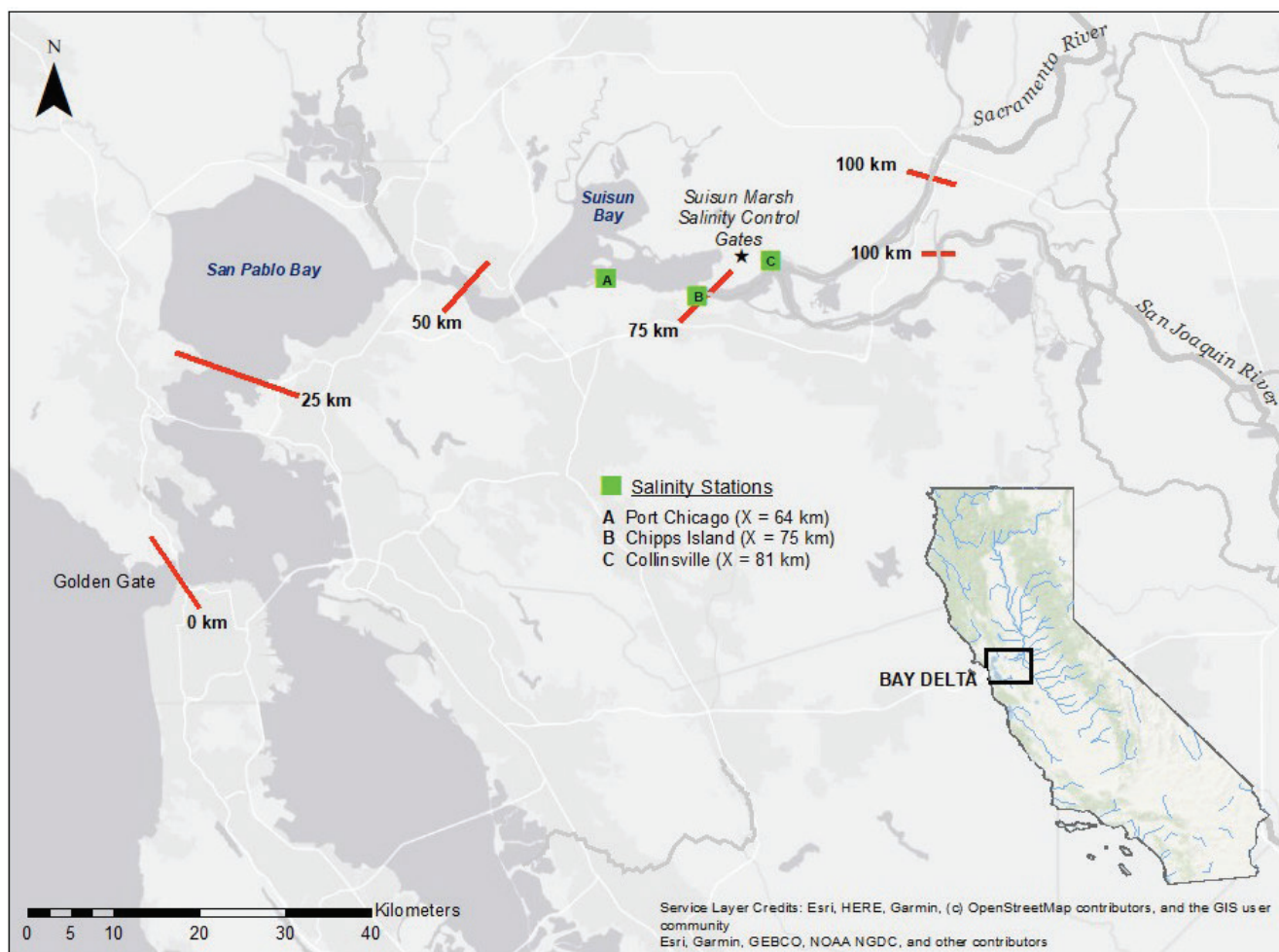


Figure 1 Map of the San Francisco Estuary

assume an average isohaline value when X_2 is located upstream of the river confluence.

Salinity Intrusion Modeling

Because of sustained interest in salinity intrusion in the Delta, a variety of modeling tools have been developed and applied since the 1980s. Our subject here is one-dimensional (1-D) empirical models. Another category of tools includes numerical (1-, 2-, and 3-D) hydrodynamic and water-quality transport models. These process-based models have been successfully used in numerous applications, ranging from operations and facility planning (CDWR 2020) to scientific exploration of fundamental estuarine mechanics (Cheng et al. 1993; Gross et al. 1999; Chua and Fringer 2011; Ateljevich et al. 2014; MacWilliams et al. 2015, 2016; Martyr-Koller et al. 2017).

Although theoretically rigorous and capable of providing insight into the basic physics of freshwater-saltwater mixing in the estuary, the data and computational requirements of these tools—especially 3-D models—often limit application for studies that require consideration of extensive (i.e., multi-decade) hydrologic sequences. Also, given that none of these process-based models have been exercised over the full range of historical hydrologic and geometric conditions, they are of unknown reliability in simulating extreme low flow conditions that occurred in the early part the 20th century. Therefore, empirical models remain useful for analysis of hypothesis testing of long hydrologic sequences as well as sequences that fall outside of typical calibration ranges and which may be

preferable under certain budget and schedule constraints.

Empirical Models for X2

This work focuses on an evaluation of six empirical X2 models (see Table 1) published since 1995. Following the conceptual model of freshwater-saltwater mixing in estuaries previously described, all these models are driven by the time history of Delta outflow in some form. Figure 2 presents a simplified diagram that shows common elements of the empirical models. These model frameworks, the data used for calibration, and the interpolation methodologies are introduced below in chronological order.

Model 1: Jassby et al. (1995)

This model, the seminal empirical X2 model for the estuary, was formulated as an autoregressive equation (Schubel 1993) and is commonly referred to as the Kimmerer-Monismith equation after its developers. Estimates of historical X2 position from this widely used model are reported and updated in CDWR's Dayflow model. Jassby et al. (1995) was originally calibrated using X2 values

interpolated from surface salinity data and outflow estimates from the Dayflow model (with modified estimates of water consumption in the Delta) from October 1967 to November 1991—the most complete data set available at the time of publication in 1995:

$$X2(t) = a + b \cdot X2(t - 1) + c \cdot \log_{10} Q(t) \quad (1)$$

where $Q(t)$ is Delta outflow and $X2(t-1)$ is the previous isohaline position expressed as distance from the Golden Gate, and a , b , and c are fitted constants. The 2 ppt bottom salinity target was converted to an equivalent surface salinity target of 1.76 ppt for X2, assuming this ratio was fixed under all flow conditions, as noted above. Salinity values from fixed locations were linearly interpolated to obtain the distance corresponding to the X2 position. Several variations of the original equation coefficients have been reported, as summarized in Bernstein (2012). In more recent work (Roy et al. 2014), the model was recalibrated for each river branch using data for several different time-periods. Because a static

Table 1 Summary of empirical X2 models for the San Francisco Estuary

Empirical model	Model no.	Independent variables ^a	No. of fitting parameters	Formulation comments	Numerical limitations
Jassby et al. 1995	1	$Q(t)$; $X2(t-1)$	3	Seminal work utilizing an autoregressive formulation	Undefined at low outflow ($Q(t) \leq 0$)
Monismith et al. 2002	2	$Q(t)$; $X2(t-1)$	3	Modified Model 1 formulation based on argument that theoretical predictions for salinity intrusion involve power-law relations rather than logarithms	Undefined at low outflow ($Q(t) \leq 0$)
MacWilliams et al. 2015	3	$Q(t)$; $X2(t-1)$	5	Modified Model 2 formulation to allow time constant for daily X2 changes to vary with outflow	Undefined at low outflow ($Q(t) \leq 0$)
Hutton et al. 2015	4	$G(t)^b$	3	Adapted antecedent outflow approach (Denton 1993) to X2 prediction such that model is defined under low outflow conditions as frequently observed under early 20th-century conditions	None
Monismith 2017	5	$Q(t)$; $X2(t-1)$	3 ^c	Formulation based on dynamic behavior of salt and velocity fields, albeit represented in simplified form	Unstable at low antecedent X2
Rath et al. 2017	6	$G(t)$; coastal water level and tidal range at Golden Gate	N/A ^d	"Hybrid" formulation, modifying Model 4 by integration with a Bayesian artificial neural network (ANN) model and incorporation of additional tidal inputs	None

a. $Q(t)$ = Delta outflow; $X2(t-1)$ = previous position of X2 isohaline; $G(t)$ = antecedent outflow.

b. Formulation can be algebraically manipulated to become a function of $Q(t)$ and $X2(t-1)$.

c. The original publication presents Model 5 with one fitting parameter. However, in our work, we refit three constants to improve predictive performance over the period of record.

d. The ANN-based formulation has many empirically determined parameters and is not directly comparable to algebraic model formulations.

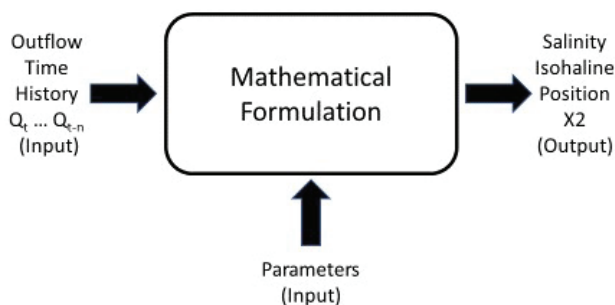


Figure 2 Simplified diagram showing common elements of the empirical X2 models

nonlinearity is applied to the model input (Q) and the model is otherwise linear, the time-series may be regarded to be of the Hammerstein type of block-structured model as described by Billings (2013). In particular, the logarithmic term in Equation 1 precludes its direct use when extremely low outflow conditions are being examined, and potentially introduces a mean bias under some flow conditions. The Jassby et al. (1995) model is hereafter referred to as “Model 1” for brevity.

Model 2: Monismith et al. (2002)

This model is conceptually similar to Model 1 described above, but the authors argue on theoretical grounds that power law relationships are superior to a logarithmic dependence on flow:

$$X2(t) = a \cdot X2(t-1) + b \cdot Q(t)^c \quad (2)$$

where a , b , and c are fitted constants. This model was originally calibrated with the same data used in Model 1.

To illustrate the assertion that these empirical models are driven by the time history of Delta outflow, Equation 2 is expanded to assist in evaluating the salinity “memory” of the Delta, where n is the number of antecedent time-steps evaluated:

$$X2(t) = a^n \cdot X2(t-n) + b \cdot \sum_{i=1}^n [a^{i-1} \cdot Q(t-i+1)^c] \quad (3)$$

By substituting typical values into the above expansion (not shown here for brevity), one can

easily show that the antecedent term $X2(t-n)$ has little influence on current salinity $X2(t)$ beyond approximately 90 days. Put another way, the value of $X2(t)$ is effectively resolved by the time history of Delta outflow over the preceding 3 months. As with Model 1, this model’s formulation precludes direct use under extremely low outflow conditions, and the model’s nonlinearity is of the Hammerstein type. The Monismith et al. (2002) model is hereafter referred to as “Model 2.”

Model 3: MacWilliams et al. (2015)

Following an approach reported by Gross et al. (2009), this model (MacWilliams et al. 2015) was calibrated with synthetic bottom salinity X2 values and Delta outflow data as produced by a 3-D hydrodynamic model of the estuary. The model was calibrated over a 3-year (April 1994–March 1997) simulation period. The conceptual basis of this model is a dynamic weighting between the two terms in Model 2:

$$X2(t) = \alpha(t) \cdot X2(t-1) + (1 - \alpha(t)) \cdot c \cdot Q^d \quad (4)$$

where $\alpha(t)$ is a dynamic function of outflow: $\alpha(t) = \alpha_0 \cdot (m \cdot Q(t) + b)$ that is bounded between 0 and 1 and α_0 , m , b , c and d are fitted constants. This work relates the dynamic $\alpha(t)$ to the flow dependence of estuarine response time described by Lerczak et al (2009), but the development is intentionally agnostic as to causes of salinity adjustment, which may include stirring and dispersion processes. Noting the intuitive appeal of relating response time with flow after MacCready (1999), Monismith et al. (2002) discussed a potential refinement to their empirical X2 formulation that was later adopted by MacWilliams et al. (2015). Assuming a linear relationship between $X2(t-1)$ and $Q(t)$, Monismith et al. (2002) found that the response time varied between 7 days at the highest flows and approximately 11 days at the lowest flows; however, they found that “... this more complicated model did not improve the fit to [surface salinity] observations nor did it reduce autocorrelation of the residuals.” Like Model 2, the power formulation in MacWilliams et al. (2015) precludes its direct use under extremely

low outflow conditions that occurred during some periods of the historical record. We note that, by substituting an antecedent outflow term for $Q(t)$, the model's time-scale of change would be related to X2 instead of instantaneous flow, and would be conceptually similar to models reported by Hutton et al. (2015) and Monismith (2017). The MacWilliams et al. (2015) model is hereafter referred to as "Model 3."

Model 4: Hutton et al. (2015)

In contrast to the autoregressive formulations adopted by the previously introduced models, Hutton et al. (2015) does not explicitly utilize $X2(t-1)$ as an independent model variable. Rather, it transforms the outflow variable into an antecedent flow, $G(t)$, a term that encodes the time history of outflow into each day's value (Denton 1993). The model, a steady state solution of Equation 2 noted in Monismith et al. (2002)—see their Equation (10)—is formulated as a simple power function of antecedent outflow:

$$X2(t) = a \cdot G(t)^b \quad (5)$$

where a and b are fitted constants, and antecedent outflow is defined by the following routing function similar to one proposed by Harder (1977):

$$\frac{\delta G}{\delta t} = \frac{(Q(t) - G(t)) \cdot G(t)}{\beta} \quad (6)$$

where $Q(t)$ is Delta outflow and β is a fitting constant with units of flow · time. Denton (1993) observed that the term $\beta/G(t)$ is a time constant that governs the rate at which $G(t)$ approaches steady state. Algebraic manipulation of Equations 5 and 6 leads to an expression in terms of Delta outflow, $Q(t)$, and antecedent salinity $X2(t-1)$:

$$X2(t) = a \cdot \left\{ \frac{Q(t)}{1 + \left[Q(t) \cdot \left(\frac{a}{X2(t-1)} \right)^{\frac{1}{b}} - 1 \right] \cdot \exp \frac{-Q(t)}{\omega_0}} \right\}^b \quad (7)$$

where $\omega_0 = \frac{\beta}{a}$. The full Hutton et al. (2015) model includes an accompanying estimate of salinity at

any position along the channel, not just the X2 isohaline. This model was originally calibrated with X2 values interpolated (log-linear) from 2000 through 2009 surface salinity data (assuming equivalence to 2.64 mS cm^{-1} specific conductance) and outflow values from the Dayflow model. In contrast to the aforementioned models, this model is well behaved under negative outflow. The Hutton et al. (2015) model is hereafter referred to as "Model 4."

Model 5: Monismith (2017)

Monismith (2017), which was derived by the author from first principles through integration of the tidally averaged salinity balance equation, adopted simplifying assumptions to yield the following autoregressive form:

$$X2(t) = v \cdot \left(\left(\frac{X2(t-1)}{L_0} \right)^{-(n+1)} - \frac{Q(t)}{Q_0} \right) + X2(t-1) \quad (8)$$

where v is a fitted constant, L_0 is a reference intrusion length, Q_0 is a reference outflow, and $n=5$. The model is fully dynamic for intrusion length, rather than being derived from steady-state relationships with lagged or integration terms added to accommodate dynamics. It incorporates work by Chen (2015) that suggests the response rate of intrusion length is proportional to the salinity state of the estuary (rather than flow) and should respond faster to outflow events when X2 is low.

The model's original calibration was based on observed bottom salinity measurements from five locations in the estuary, along with corresponding Dayflow-based outflow data, for a 9-month period that spanned October 2014 through June 2015, which is a period of extreme drought punctuated by high flow. X2 values were estimated from observed bottom salinity data that were low-pass filtered with a Godin filter then smoothed spatially with a spline interpolation. Monismith (2017) also speculatively manipulated the training outflow time-series, reducing it by 2000 cfs. The sensitivity of the empirical X2 models to such a manipulation is described later in this paper. The Monismith (2017) model is hereafter referred to as "Model 5."

Model 6: Rath et al. (2017)

This model employs machine learning techniques to improve upon the X2 estimates produced by a reference model (Model 4). Artificial neural networks (ANNs), trained on several flow moving averages and tidal variables (water level and tidal range at Golden Gate), were fit to the residuals between observed X2 and predictions from the reference model:

$$X2(t) = a \cdot G(t)^b + \varepsilon \quad (9)$$

where ε is the reference model residual. Two ANNs were developed to distinguish between the Sacramento and San Joaquin rivers' branches of the estuary. For this work, X2 predictions from Rath et al. (2017) are reported as the average value from the two ANN model outputs. The model, which uses a more complex fitting algorithm and additional inputs besides Delta outflow, is not as easily implemented as the aforementioned models. The model was included in this survey to provide an empirical point of comparison with the other five empirical models on skill. The Rath et al. (2017) model is hereafter referred to as "Model 6."

METHODS

The methods used in this survey of X2 isohaline empirical models of the estuary were founded on the development and assembly of outflow and salinity data sets spanning nearly a century (WYs 1922 through 2017), and recalibration of the empirical models described above (excepting Model 6) to a common period. These recalibrated models were then subjected to a variety of statistical and perturbation tests of model performance. We used the more complex ANN-based Model 6 to compare the performance of models but did not recalibrate it. Details associated with the survey methodology are summarized below.

Data

Net freshwater flows out of the Delta are difficult to measure directly. As a result, several estimation approaches have been proposed

over the years (CDWR 2016; Fleenor et al. 2016). While advances in direct flow measurements continue to occur, the regulatory definition of Delta outflow continues to be quantified through a water balance procedure known as the "Net Delta Outflow Index" (NDOI). NDOI—a quantity that represents daily average flow at the confluence of the Sacramento and San Joaquin rivers—is calculated as the sum of daily Delta river inflows along its periphery, minus net Delta channel depletions and Delta exports, and is reported in the Dayflow model. Monismith (2016) examined differences between NDOI and direct flow measurements over a recent historical interval (WYs 2008 through 2014) and found the two measures were coherent as time-series, although he observed that the root mean square difference between the two values is comparable to the magnitude of typical low flow NDOI values. Any systematic flaws in outflow will affect the calibration of parameters and the long-term accuracy of both empirical and mechanistic models, particularly in the upper estuary. For our work, NDOI was used as an estimate for daily outflow for the period spanning WYs 1930 through 2017. We assembled daily outflow before October 1929 from work presented in Hutton et al. (2015).

An extensive data synthesis and cleaning effort reported in Hutton et al. (2015) that spanned WYs 1922 through 2012 provided most of the salinity data and X2 estimates used for our study. As summarized in the following paragraph, we updated this data set through WY 2017 as part of this work. Hutton et al. (2015) queried databases with continuous conductivity data from several sources, including the California Data Exchange Center, the Interagency Ecological Program and the US Environmental Protection Agency's STORET. These data were further supplemented by US Geological Survey (USGS) data to represent high outflow periods when the low salinity zone extended far downstream into San Francisco Bay. The Hutton et al. (2015) procedures resulted in a master database with relatively complete conductivity-based records for the sub-period that spanned WYs 1968 through 2012; this sub-period represents the system after completion of

major surface water storage and export projects in and upstream of the Delta. Some stations had a conductivity record before October 1967, but the coverage of stations across the gradient was incomplete until WY 1968. Hutton et al. (2015) also assembled legacy grab sample data across the estuary to develop a daily X2 record for the earlier sub-period that spanned WYs 1922 through 1967.

We used the same procedures to extend the data set through WY 2017. We downloaded the latest flow and salinity data from CDEC and USGS to update the interpolated X2 data set from October 2012 through September 2017. We downloaded these updated data from the stations along the estuary, and they are summarized in the supplemental information (Table A1 in Appendix A). For consistency with other published work, we used an average isohaline value when X2 is located upstream of the confluence of the Sacramento and San Joaquin rivers. Our analysis focused on the more recent period that spanned WYs 1968 through 2017; however, a limited evaluation also considered the full period of record back to WY 1922.

Model Recalibration

We recalibrated the empirical X2 models identified above (excluding Model 6) with common salinity (X2) and outflow (NDOI) data sets from WYs 2000 through 2009. We selected the calibration period to span a wide range of flow conditions that have been observed since construction of major reservoir and Delta export facilities. The 10-year period includes 1 wet water year, 4 above/below water years, and 5 dry/critical water years. Although the original calibration period of Model 4 aligned with our study's recalibration period, for consistency with the other models we recalibrated it to X2 values that represented average positions along the Sacramento and San Joaquin river branches. We recalibrated all models with X2 values interpolated from near-surface salinity observations (we also refer to this as "observed" X2) using a consistent log-linear interpolation methodology. Recalibration for all models used a Markov Chain Monte Carlo (MCMC) procedure applied to a Bayesian normal likelihood. The prior

distributions used in the inference procedure were informative but very weak relative to the amount of training data. Model 3 was originally fit using a Bayesian MCMC procedure, and for this work we used priors similar to those reported in MacWilliams et al. (2015). Details on model recalibration are provided in the supplemental information (Appendix A).

Statistical and Sensitivity Analyses

We tested and compared empirical X2 model performance in several ways using statistical and sensitivity analyses. The methodologies associated with each approach are described below.

We tested model performance statistically over the 50-year period of record (WYs 1968 through 2017) using standard metrics, including the square of the correlation coefficient (R-squared), the Nash-Sutcliffe efficiency (NSE) index, the mean residual, and the RMSE. The NSE index indicates a perfect model fit for a value of 1. An NSE index approaching 0 indicates a model fit that is no more accurate than simply using the observed mean; an index value below zero indicates a model fit that is poorer than simply using the observed mean. We also used Sen's slope (Sen 1968)—a robust estimate of time trend—to evaluate potential model residual trends over the full period of record (WYs 1922 through 2017).

We partitioned X2 and outflow data in several ways to conduct the statistical tests for the 50-year period of record, including three bins that represented calibration (WYs 2000 through 2009) and validation (WYs 1968 through 1999 and 2010 through 2017) periods, 12 bins that represented months, 50 bins that represented individual water years, and 9 bins that represented different hydrologic ranges as measured by current day outflow and antecedent (previous-day) isohaline position $X2(t-1)$. The empirical X2 models tend to perform best under conditions in the neighborhood of equilibrium, when antecedent X2 and outflow are inversely related. Conditions that are inconsistent with this relationship (e.g., a high outflow event after an extended period of high salinity intrusion) tend to be poorly replicated

Table 2 Number of daily outflow and X2 data points (WYs 1968–2017) by hydrologic ranges as defined by current day outflow $Q(t)$ and antecedent (previous-day) isohaline position $X2(t-1)$

Bin	$Q(t) < 5,000$ cfs		$5,000 < Q(t) < 30,000$ cfs		$Q(t) > 30,000$ cfs	
	<i>N</i>	Bin no.	<i>N</i>	Bin no.	<i>N</i>	Bin no.
$X2(t-1) < 65$ km	1	11	1,131	12	2,971	13
$65 \text{ km} < X2(t-1) < 85$ km	1,038	21	6,715	22	515	23
$X2(t-1) > 85$ km	3,538	31	1,901	32	53	33

by the empirical X2 models, and reflect highly dynamic conditions not easily encapsulated in a single outflow time history. To differentiate empirical X2 model performance accordingly, we used a 9-bin data partition (as cited above). The number of data points in each bin are presented in Table 2.

After the statistical analyses, we explored steady state and dynamic model behaviors through a variety of sensitivity tests. First, we evaluated steady state flow at three characteristic X2 positions in the estuary that are of regulatory importance. Next, we tested dynamic response by evaluating the immediate (single-day) rates of change in intrusion length excited by various changes in flow in a neighborhood around steady state. We further confirmed time-scales of change by evaluating dynamic model response to a step change in flow as reported in Monismith (2017). Specifically, we tested how long it took each model to make each possible one-step transition (high to mid, mid to low, and vice versa) between the three characteristic positions under a step change in steady-state flows. We further tested dynamic model response by imposing a 14-day period outflow oscillation around three equilibrium conditions.

Given the regulatory importance associated with quantifying Delta outflow, and given the aforementioned difficulties associated with directly measuring net flows in the highly tidal environment around the Sacramento and San Joaquin rivers' confluence, some have proposed the use of salinity data to infer outflow (CDWR 2016; Fleenor et al. 2016). Implicit in this proposal is the assumption that a given salinity and salinity

gradient tightly constrain the outflow history. To test this assumption, we again recalibrated the empirical X2 models, but this time with perturbed outflow time-series. Specifically, we adjusted the Dayflow outflow time-series uniformly upward and downward by 2,000 cfs, similar to a change reported by Monismith (2017). Thus, this analysis compared the frequency of model residuals from three distinct outflow scenarios: baseline NDOI (the initial recalibration), increased NDOI, and decreased NDOI. We compared residual time-series from these scenarios to see if the contrived outflow modifications produced poorer model fits. We hypothesize that poorer fits would indicate a tightly constrained relationship between salinity and outflow, and lend support to the use of salinity measurements to infer Delta outflow. Model 6 was not subjected to these sensitivity analyses because similar flow conditions were not incorporated in its training data set.

RESULTS

Results from this survey of X2 isohaline empirical models of the estuary are presented below, following the methodology outlined above.

Model Recalibration

The original and recalibrated parameter values for each of the empirical X2 models (excluding Model 6) are shown in Table 3. Recalibration to a common data set generally modified the original published values of the fitting parameter values; however, the changes are not dramatic. Although Model 5 constants Q_0 , L_0 , and n were not presented by the author as fitting constants, we refit the first two constants as part of this

Table 3 Original and revised empirical X2 model parameters after recalibration. Units are consistent with original publications; all are reported in metric units except Model 4 (which reports in mixed metric and English units). Recalibrated parameter values shown here are posterior means.

Model 1: Jassby et al. (1995)

$X2(t) = a + b \cdot X2(t-1) + c \cdot \log_{10} Q(t)$	<i>a</i> (km)	<i>b</i>	<i>c</i> (km m ⁻³ s)
Original	10.2	0.945	-2.30
Recalibrated	12.8	0.917	-2.60

Model 2: Monismith et al. (2002)

$X2(t) = a \cdot X2(t-1) + b \cdot Q(t)^c$	<i>a</i>	<i>b</i> (km m ⁻³ s)	<i>c</i>
Original	0.919	13.57	-0.141
Recalibrated	0.919	17.50	-0.186

Model 3: MacWilliams et al. (2015)

$X2(t) = \alpha(t) \cdot X2(t-1) + (1 - \alpha(t)) \cdot c \cdot Q^d$ $\alpha(t) = \min[1, \max[0, \alpha_0 \cdot (m \cdot Q(t) + b)]]$	<i>c</i> (km m ⁻³ s)	<i>d</i>	α_0	<i>m</i> (m ⁻³ s)	<i>b</i>
Original	284	-0.23	1.025	-1.12 x 10 ⁻⁴	0.977
Recalibrated	241	-0.20	1.018	-1.12 x 10 ⁻⁴	0.977

Model 4: Hutton et al. (2015)

$X2(t) = a \cdot G(Q(t), \beta)^b$	<i>a</i> (km * cfs ^{-b})	<i>b</i>	β (cfs-years)
Original (Sacramento River)	456	-0.193	475
Original (San Joaquin River)	502	-0.203	475
Recalibrated (Average)	473	-0.197	475

Model 5: Monismith (2017)

$X2(t) = v \cdot \left(\frac{X2(t-1)}{L_0} \right)^{-(n+1)} - \frac{Q(t)}{Q_0} + X2(t-1)$	<i>v</i> (km)	Q_0 (m ³ s ⁻¹)	L_0 (km)	<i>n</i>
Original	0.50	141	85	5
Recalibrated	0.31	143	87	5

work to improve predictive performance over the period of record. Figure 3 shows an illustrative time-series result for a specific year (WY 2008) comparing observed X2 and modeled X2 values from the updated calibrations. The bottom panel shows the NDOI value for each day on a log scale to compare salinity behavior with its primary driver. The models generally capture observed salinity behavior during this representative calibration period, although prediction across models varies from 5 to 10 km. For this period, none of the recalibrated models appear to fully

capture the high flow winter events. Across other calibration years (data not shown), we similarly observe that general patterns are reasonably replicated, but more extreme events are not reproduced as well.

Model 1, formulated as a logarithmic relationship, is undefined when outflow is less than or equal to zero. Similarly, Models 2 and 3 are formulated as power-law relationships, and are generally undefined when outflow is less than or equal to zero. Extreme salinity intrusion events

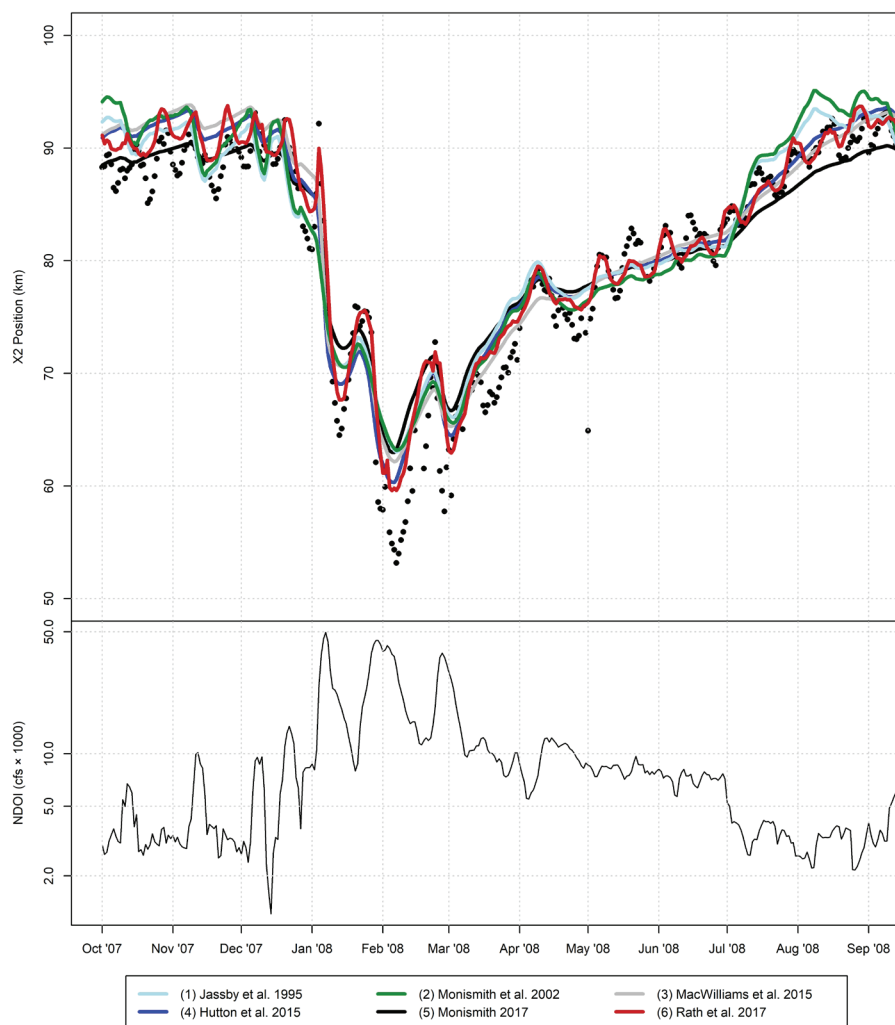


Figure 3 Time-series of observed and predicted X2 (upper panel) and Delta outflow (lower flow) for a representative year. Model predictions are based on recalibrations conducted as part of this work.

associated with persistent negative net outflow conditions are seen in the historical NDOI record, particularly before construction of Shasta Reservoir in 1944. Under such conditions, these models are incapable of producing a complete time-series of salinity predictions. In Schubel (1993), Kimmerer and Monismith recommend constraining the outflow record to a minimum value of 316 cfs when applying Model 1. In Bernstein (2012), challenges associated with predicting salinity under extreme salinity intrusion were noted; as a remedy, the outflow record was constrained to a minimum value of 50 cfs when applying the same model. Through sensitivity analysis, we found that imposing a substantially higher minimum outflow constraint of 1,800 cfs minimized model residuals; we

imposed this constraint on the NDOI time-series as part of the full period (WYs 1922 through 2017) evaluation process of Models 1, 2, and 3. As part of the recalibration process, these models were reinitialized with observed X2 positions when predictions were undefined.

As part of this work, we discovered that Model 5 is subject to oscillatory behavior under extremely high outflow conditions, resulting in alternating sequences of large and small X2 estimates during the high flow event. Specifically, the term $\left(\frac{X2(t-1)}{L_0}\right)^{-(n+1)}$ in Equation 8 becomes very large when the numerator $X2(t-1)$ is small compared to the denominator (i.e., reference isohaline length (L_0) with a recalibrated value of 87 km) and dominates

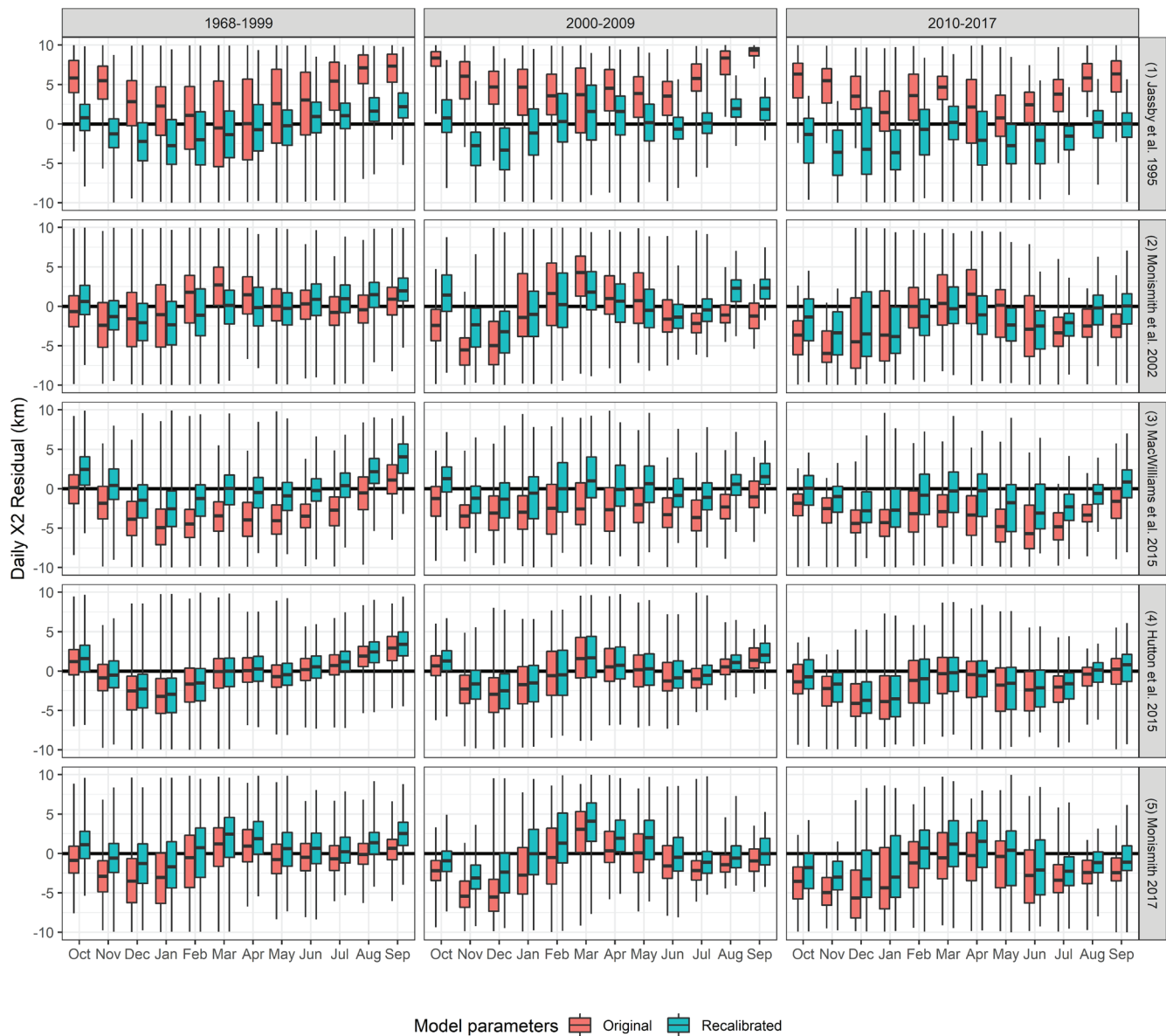


Figure 4 Comparison of daily X2 model residuals under the originally published parameters and the recalibrated parameters used in the rest of this work. Residual distributions are shown as boxplots for three different WY time-periods: 1968–1999 (**left**), 2000–2009 (**center**, the calibration period for recalibration), and 2010–2017 (**right**). Each combination of model and time-period then shows the daily residual distribution for each month within that category.

the term $\frac{Q(t)}{Q_0}$. To censure the relatively few data points associated with these extremely high outflow conditions, a minimum constraint of 45 km was imposed on model output as part of the model’s recalibration and evaluation process.

Using the 50-year surface salinity data set (WYs 1968 through 2017), we compared the recalibrated model residuals on a monthly basis with those

obtained using the original models (i.e., using the parameter values exactly as published) (Figure 4). The comparisons are shown for WYs 1968 through 1999, WYs 2000 through 2009 (the calibration period), and WYs 2010 through 2017 (a period including some extreme drought years in 2014-2015; this also included a year, 2015, where a salinity barrier was installed to reduce salinity intrusion). This is a broader comparison

than reported in the original publications. An important observation from this multi-decade comparison is that, with some exceptions, the original and recalibrated models perform similarly, thus confirming that we have not fundamentally altered model behavior through our recalibration process. A closer review of the plots shows that the performance of the recalibrated models is generally better than or similar to the original models (residuals closer to zero indicate better performance). Henceforth, results are presented only for the recalibrated models.

Statistical Analyses

Water Years 1968 through 2017

Formal statistical comparison of empirical X2 model performance over different time-periods is reported in tabular form using different metrics. Table 4 summarizes model statistics for three different sub-periods of the 50-year data record that spans WYs 1968 through 2017. As discussed earlier, the preceding sub-period (WYs 1968 through 1999) encompasses the multi-year

drought of WYs 1987 through 1992, WYs 2000 through 2009 represent the model recalibration period, and the following sub-period (WYs 2010 through 2017) encompasses the recent multi-year drought of WYs 2012 through 2016. Performance is reasonably strong across all models as measured by the reported metrics. The recent period of WYs 2010 through 2017 is associated with somewhat worse model performance relative to the calibration period; the recent period is more challenging to replicate according to the binning scheme of Table 2: outflow and salinity data fall in Bins 13, 22, and 31 (representing quasi-steady state conditions) an average of 142 days per year during the calibration period; whereas, the data fall in the same bins only 124 days per year during the recent period. This recent period is characterized by more dynamic conditions and includes more days of very high flow (Bin 13) and high X2 despite medium outflow (Bin 32). Model performance by month is presented in the supplementary information (Table A2).

Table 4 Model performance statistics across three sub-periods of the 50-year time-series spanning WYs 1968–2017

Period (WYs)	Model	Model no.	N	R ²	NSE	Mean residual (km)	RMSE (km)
1968–1999	Jassby et al. 1995	1	11375	0.92	0.91	-0.29	4.04
	Monismith et al. 2002	2	11375	0.92	0.91	0.22	4.17
	MacWilliams et al. 2015	3	11375	0.94	0.93	0.10	3.54
	Hutton et al. 2015	4	11394	0.94	0.94	0.05	3.46
	Monismith 2017	5	11393	0.94	0.93	0.60	3.51
	Rath et al. 2017	6	11394	0.96	0.96	0.23	2.85
2000–2009	Jassby et al. 1995	1	3586	0.90	0.90	0.00	3.85
	Monismith et al. 2002	2	3586	0.89	0.89	0.01	3.97
	MacWilliams et al. 2015	3	3586	0.92	0.92	0.01	3.39
	Hutton et al. 2015	4	3589	0.92	0.92	0.01	3.33
	Monismith, 2017	5	3589	0.93	0.90	0.19	3.72
	Rath et al. 2017	6	3589	0.95	0.94	0.28	2.87
2010–2017	Jassby et al. 1995	1	2880	0.90	0.88	-1.82	4.55
	Monismith et al. 2002	2	2880	0.90	0.88	-1.80	4.64
	MacWilliams et al. 2015	3	2880	0.93	0.91	-1.49	3.93
	Hutton et al. 2015	4	2880	0.94	0.92	-1.57	3.75
	Monismith 2017	5	2880	0.94	0.89	-1.29	4.40
	Rath et al. 2017	6	2880	0.93	0.91	-1.27	4.00

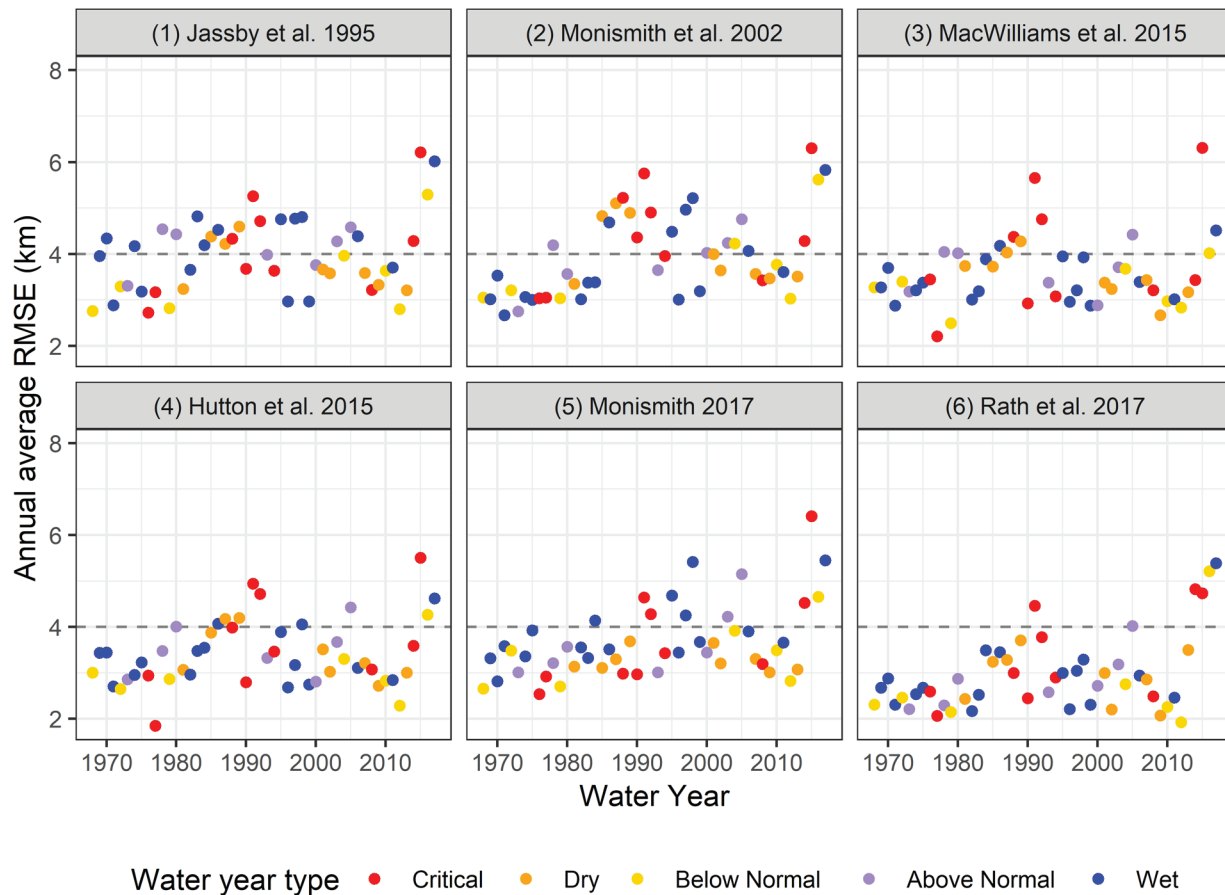


Figure 5 Annual average root mean square error (RMSE) for recalibrated X2 models for WYs 1968–2017. Each point represents the mean of the square of the daily residuals for a given model and water year. Points are colored according to the water year classification derived from the Sacramento Valley Index.

Figure 5 shows the annual average RMSE for each model over the 50-year data record. Data points are color-coded according to the Sacramento Valley Index water year classification. Some consistent trends are visually apparent. For example, WY 2015 records the highest RMSE for all models, coincident with extreme drought conditions and placement of a temporary rock barrier in the western Delta to manage salinity (Lund et al. 2018). Another commonly observed trend (see Models 1, 2, and 5) is relatively high model residuals coincident with the extended drought of WYs 1987 through 1992. Empirical X2 model performance does not appear to depend entirely on water year type: although the most poorly simulated years tend to be critical years,

poor model performance is not universally associated with critical conditions.

Table 5 shows model statistics partitioned in 9 bins according to antecedent salinity (previous-day X2 position) and NDOI as defined in Table 2. As previously discussed under Methods, examining model performance in this manner is helpful because these models tend to be most accurate under quasi-steady state conditions. When antecedent salinity is consistent with the prevailing outflow regime, one can be reasonably confident that outflow conditions have not changed drastically over a very short time interval (e.g., a sudden storm event), and the notion of a daily average X2 is close to a single position in the estuary. The influence of quasi-

Table 5 Model skill statistics across nine data bins that represent different hydrologic ranges over the 50-year time-series spanning WYs 1968–2017. The data bins are defined by current day outflow $Q(t)$ and antecedent (previous day) isohaline position $X2(t-1)$.

Bin no.	$X2(t-1)$ Bin	$Q(t)$ Bin	Model	Model no.	N	Mean residual (km)	RMSE (km)
12	$X2(t-1) < 65$ km	$5,000 \text{ cfs} < Q < 30,000 \text{ cfs}$	Jassby et al. 1995	1	1131	0.94	3.91
			Monismith et al. 2002	2	1131	0.78	3.61
			MacWilliams et al. 2015	3	1131	1.71	3.55
			Hutton et al. 2015	4	1131	1.76	3.38
			Monismith 2017	5	1131	4.12	4.94
			Rath et al. 2017	6	1131	1.47	3.07
13	$X2(t-1) < 65$ km	$Q > 30,000 \text{ cfs}$	Jassby et al. 1995	1	2971	-1.46	5.09
			Monismith et al. 2002	2	2971	0.83	4.52
			MacWilliams et al. 2015	3	2971	0.17	3.42
			Hutton et al. 2015	4	2971	-0.40	3.39
			Monismith 2017	5	2971	3.03	4.60
			Rath et al. 2017	6	2971	0.33	3.07
21	$65 \text{ km} < X2(t-1) < 85$ km	$Q < 5,000 \text{ cfs}$	Jassby et al. 1995	1	1026	3.00	3.95
			Monismith et al. 2002	2	1026	3.01	4.25
			MacWilliams et al. 2015	3	1026	-0.08	2.88
			Hutton et al. 2015	4	1038	1.03	2.55
			Monismith 2017	5	1038	0.18	2.09
			Rath et al. 2017	6	1038	0.53	2.19
22	$65 \text{ km} < X2(t-1) < 85$ km	$5,000 \text{ cfs} < Q < 30,000 \text{ cfs}$	Jassby et al. 1995	1	6714	0.07	2.95
			Monismith et al. 2002	2	6714	-0.73	3.03
			MacWilliams et al. 2015	3	6714	-0.05	3.27
			Hutton et al. 2015	4	6715	0.02	2.88
			Monismith 2017	5	6714	0.73	2.85
			Rath et al. 2017	6	6715	0.42	2.61
23	$65 \text{ km} < X2(t-1) < 85$ km	$Q > 30,000 \text{ cfs}$	Jassby et al. 1995	1	515	-2.19	4.66
			Monismith et al. 2002	2	515	-2.20	4.68
			MacWilliams et al. 2015	3	515	-3.71	5.47
			Hutton et al. 2015	4	515	-4.85	6.11
			Monismith 2017	5	515	-2.27	4.38
			Rath et al. 2017	6	515	-0.72	3.45
31	$X2(t-1) > 85$ km	$Q < 5,000 \text{ cfs}$	Jassby et al. 1995	1	3529	0.24	3.82
			Monismith et al. 2002	2	3529	1.42	4.81
			MacWilliams et al. 2015	3	3529	0.51	3.48
			Hutton et al. 2015	4	3538	0.62	3.44
			Monismith 2017	5	3538	-1.74	3.25
			Rath et al. 2017	6	3538	-0.38	3.35

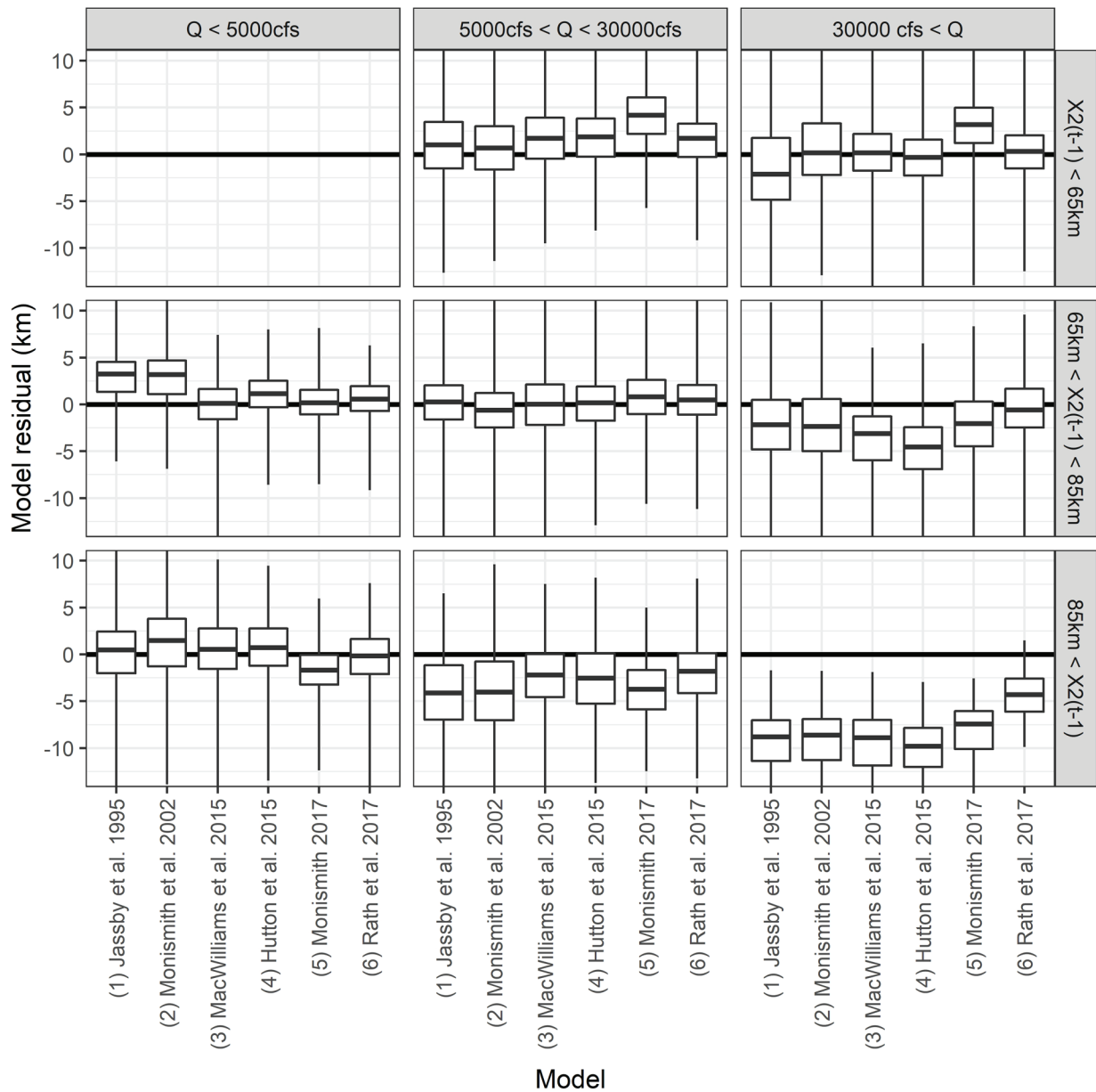


Figure 6 Boxplot of model residuals (predicted–observed) for the recalibrated empirical X2 models over the 50-year time-series spanning WYs 1968–2017. The data are partitioned by hydrologic ranges as defined by current day outflow and previous day isohaline position. Plot range is restricted to –10 km to 10 km for visual clarity, censoring only a very small number of points that fall outside that range. The first panel is empty because only 1 data point is associated with that bin. For each category, the *box* spans the 25th to 75th percentiles, and the *horizontal line* is the median value of the residuals. The *vertical lines* extend to the most extreme residual located less than 1.5 times the height of the box away from it. All other points are shown as *semi-transparent discrete points*.

steady state conditions on model performance is visually demonstrated in Figure 6, where the panels aligned with the diagonal that runs from bottom-left to top-right (Bins 13, 22, and 31) show model residuals more closely centered on 0 km.

Table 6 sorts the magnitude of the mean residual for each model and bin into three different subjective skill categories: excellent (0 km to 2 km), adequate (2 km to 5 km), and poor (>5 km). The span of the first category is meant

Table 6 Mean residual performance statistics from Table 5 sorted into three subjective skill categories: excellent (0 km to 2 km) (*green*); adequate (2 km to 5 km) (*yellow*); and poor (> 5 km) (*red*). Skill categories reflect absolute values.

Bin no.	$Q(t)$ bin	$X2(t-1)$ bin	Model 1 Jassby et al. (1995)	Model 2 Monismith et al. (2002)	Model 3 MacWilliams et al. (2015)	Model 4 Hutton et al. (2015)	Model 5 Monismith (2017)	Model 6 Rath et al. (2017)
12	5,000 cfs < $Q(t)$ < 30,000 cfs	$X2(t-1) < 65$ km	Green	Green	Green	Green	Yellow	Green
13	$Q(t) > 30,000$ cfs	$X2(t-1) < 65$ km	Green	Green	Green	Green	Yellow	Green
21	$Q(t) < 5,000$ cfs	65 km < $X2(t-1)$ < 85 km	Yellow	Yellow	Yellow	Green	Green	Green
22	5,000 cfs < $Q(t)$ < 30,000 cfs	65 km < $X2(t-1)$ < 85 km	Green	Green	Green	Green	Green	Green
23	$Q(t) > 30,000$ cfs	65 km < $X2(t-1)$ < 85 km	Yellow	Yellow	Yellow	Yellow	Yellow	Green
31	$Q(t) < 5,000$ cfs	$X2(t-1) > 85$ km	Green	Green	Green	Green	Green	Green
32	5,000 cfs < $Q(t)$ < 30,000 cfs	$X2(t-1) > 85$ km	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
33	$Q(t) > 30,000$ cfs	$X2(t-1) > 85$ km	Red	Red	Red	Red	Red	Yellow

to represent our experience with the limits of current practice to resolve daily $X2$ as a specific single value, even with the most flexible models and in well-behaved hydrodynamic conditions. All models provided excellent skill under quasi-steady conditions of “medium outflow, medium antecedent $X2$ ” (Bin 22) and “low outflow, high antecedent $X2$ conditions” (Bin 31); all but Model 5 provided excellent skill under the remaining quasi-steady condition of “high outflow, low antecedent $X2$ ” (Bin 13).

For data bins that represent less common outflow and antecedent salinity conditions (representing unsteady conditions), all but Model 5 provided excellent skill under “medium outflow, low antecedent $X2$ ” conditions (Bin 12), and all but Models 1 and 2 provided excellent skill under “low outflow, medium antecedent $X2$ ” conditions (Bin 21). All models provided adequate skill under “medium outflow, high antecedent $X2$ ” (Bin 32). For the remaining and relatively rare data bins, Model 6 was identified as the best performer by providing excellent skill under “high outflow, medium antecedent $X2$ ” conditions (Bin 23) and adequate skill under “high outflow, high antecedent $X2$ ” conditions (Bin 33).

Water Years 1922 through 2017

Prior analysis of historical data (Hutton 2014; Hutton et al. 2015) indicates little change over the past 9 decades in the accuracy or nature of flow–salinity relationships. We tested this assertion against the suite of empirical $X2$ model predictions over the full observational record that spans WYs 1922–2017.

Figure 7 depicts the time-series of monthly average model residuals extending back to WY 1922. Consistent and significant residual trends over this time-frame would indicate systematic changes in the estuary’s flow–salinity relationship that could not be explained by models as calibrated to conditions for WYs 2000–2009. The sample size is sufficiently large that a robust estimate of time trend using Sen’s slope (Sen 1968) was nominally significant for all cases. However, the estimated trends are generally no more than a few tenths of a kilometer per decade, which is small compared to the general noise scale of model predictions (which is on the order of 1 to 2 km for the best fits). Residual trends are generally negative, and typically reflect greater model variance in the early part of the record. Residual trends are positive in 14 of the 72 panels

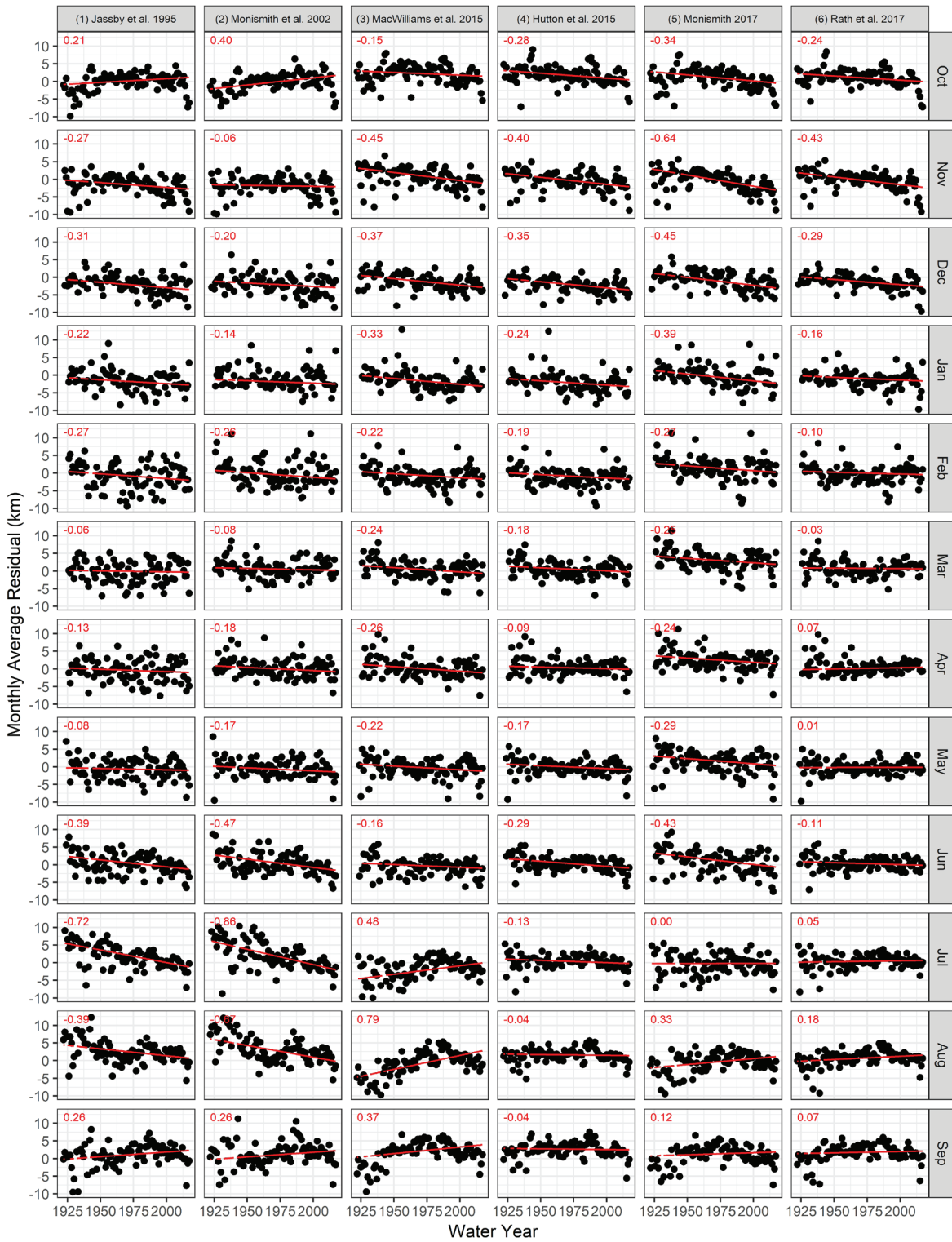


Figure 7 Trends in monthly average residuals by month for each of the recalibrated empirical X2 models over the full period of record spanning WYs 1922–2017. The *red lines* depict Theil–Sen linear trend estimates (Sen 1968). The range of the y-axis on each panel is restricted to +/- 10 km for visual clarity, even though a small number of points (occurring predominantly in the early part of the record) have larger absolute residuals that fall outside that range. *Red numbers* are the estimate of the decadal time trend in residuals.

Table 7 Steady-state outflow required to maintain X2 at a characteristic position in the San Francisco Estuary. Outflow is reported in units of cubic feet per second (cfs).

Model	Model no.	X2 = 64 km	X2 = 75 km	X2 = 81 km
Jassby et al. 1995	1	26,200	11,600	7,400
Monismith et al. 2002	2	25,100	10,700	7,000
MacWilliams et al. 2015	3	28,600	12,800	8,700
Hutton et al. 2015	4	26,400	11,800	8,000
Monismith 2017	5	32,400	12,500	7,900

displayed in [Figure 7](#); these occur primarily in the summer and fall months of July through October (all but Model 4 show positive trends in September). Downward trends of notably high magnitude are associated with Models 1 and 2 during the summer months of June, July, and August. These downward trends are associated with persistent model over-prediction in the early part of the record, and may reflect the limited ability of these models to simulate extremely low flow conditions observed in the pre-Shasta Reservoir data record (resulting from their logarithmic and power-law formulations). Other notably high downward trends are associated with Models 3, 4, 5, and 6 during the month of November. For all practical purposes, except for the low flow limitations previously identified for a subset of the models, the empirical models show no meaningful error patterns as represented by the residuals, and therefore appear adequate for representing X2 behavior over the past century.

Sensitivity Analyses

We explored steady state and dynamic behaviors of the empirical X2 models (excluding Model 6) through five sensitivity analyses. Results from these analyses are reported below.

Steady-State Relationships

[Table 7](#) shows the outflow required to maintain a steady X2 value at three characteristic isohaline positions for all empirical X2 models except Model 6. The three positions correspond to monitoring locations associated with current X2 regulations (CSWRCB 2000): Port Chicago (i.e., Roe Island) at 64 km, Mallard Island (i.e., Chippis Island) at 75 km, and Collinsville at 81 km ([Figure 1](#)). These calculations used high-precision

recalibrated model parameters (rounded values are provided in [Table 3](#)) and were adjusted analytically or manually until a steady state condition was achieved (with X2 changing by less than 0.01 km). The results demonstrate similar steady state relationships between outflow and X2, with required outflows ranging between 25,100 to 32,400 cfs to maintain X2 at 64 km, 10,700 to 12,800 cfs to maintain X2 at 75 km, and 7,000 to 8,700 cfs to maintain X2 at 81 km. The variation in required flow between models is commensurate with the uncertainty usually associated with outflow measurement. For comparison, flow-based alternatives to the X2 regulations at Roe Island, Chippis Island, and Collinsville are 29,200 cfs, 11,400 cfs, and 7,100 cfs, respectively.

Time Rates of Change Away from Equilibrium

The empirical X2 models differ more in their dynamic behavior than in their steady state or equilibrium behavior. [Figure 8](#) shows model response to a single day change in outflow for each empirical model (excluding Model 6) at five selected antecedent X2 positions: 55 km, 65 km, 75 km, 85 km, and 95 km. The antecedent X2 can be considered a reference and the perturbations the rate of change normalized to a day. The model predictions tend to converge under quasi-steady state conditions when $X2(t) \approx X2(t-1)$, with Model 5 being an exception when the previous-day X2 (i.e., antecedent X2) is further downstream at 55 km. When outflow deviates from steady conditions, however, the models begin to differ in their rate of response. Models 3, 4, and 5 generally exhibit flatter X2 responses, meaning that they do not tend to move away from steady state except under low-salinity equilibrium conditions (e.g., $X2(t-1) = 55$ km). During a freshet, however, these

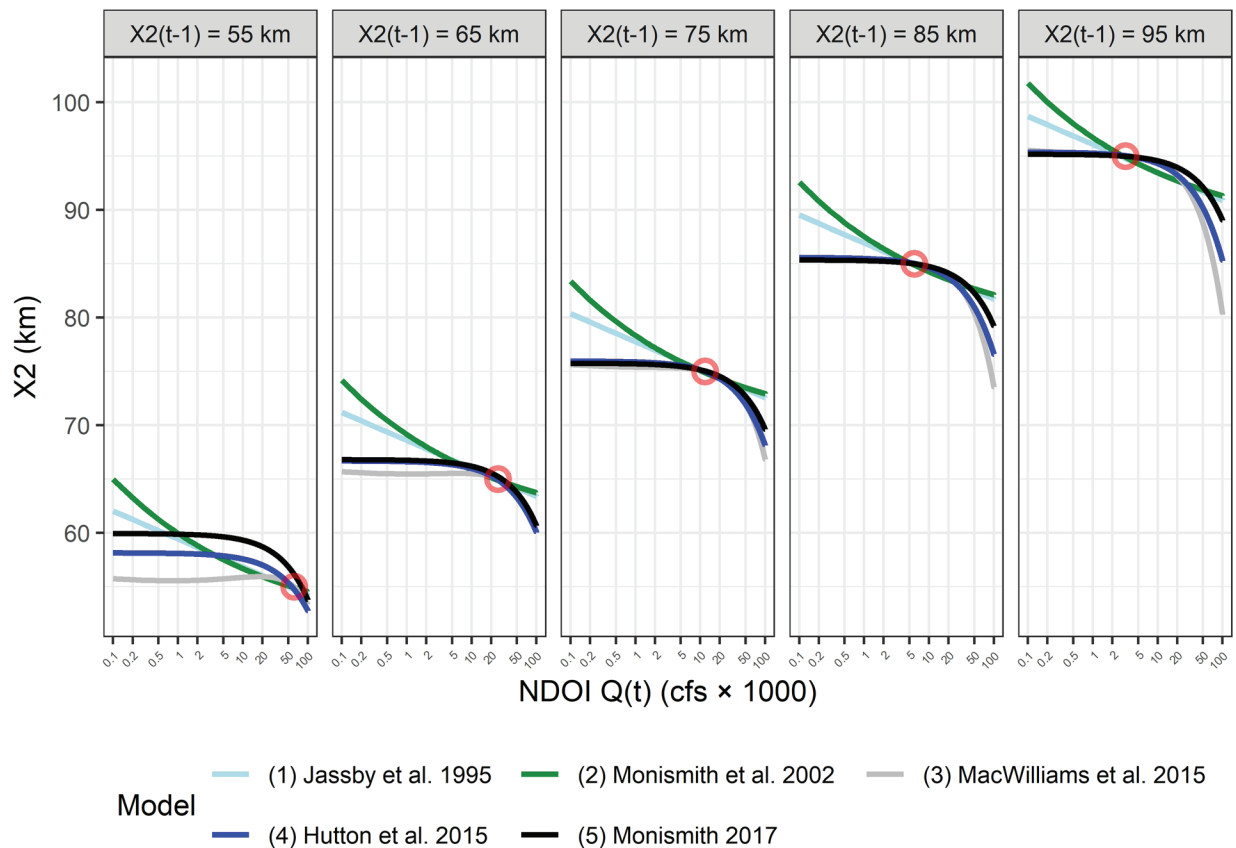


Figure 8 This figure shows the dynamic response to a 1-day change in Delta outflow associated with each empirical X_2 model (excluding Model 6) at five selected antecedent salinity conditions. For a given antecedent condition, the model predictions tend to converge under quasi-steady conditions when $X_2(t) \approx X_2(t-1)$ as highlighted by the red circle.

three models can exhibit a steep second-order effect such that they respond more—and more quickly—under large increases in flows. The remaining models share a different qualitative response that is approximately log-proportional over the full range of outflow change.

Step Response

Figure 9 shows model response to a step change in Delta outflow for each empirical X_2 model (excluding Model 6). Specifically, the figure shows the transitions between neighboring characteristic isohaline positions of 64, 75, or 81 km currently used for X_2 regulations (CSWRCB 2000). The double e-folding time (86%)—or the time taken for X_2 to change by $1/e^2$ of its initial value—is indicated with a dot; this notation is common in describing exponential decay or growth functions.

Models 1 and 2 have a transition time that is mostly independent of the state of the estuary. These models predict that it takes approximately 25 days to make any of the transitions. By contrast, Models 3, 4, and 5 strongly exhibit salinity-dependent time-scales, with values that vary from 14 to 57 days—a range which straddles the 25 days of response time of the other models. Models 3, 4, and 5 react more rapidly under low antecedent X_2 conditions (i.e., moving toward or away from a steady state position of 64 km) than they do under high antecedent X_2 conditions (i.e., moving toward or away from a steady state position of 81 km) and respond faster to decreasing (freshening) X_2 than increasing X_2 .

Fortnightly Oscillation

Fortnightly excitation is a common feature of tidal forcing in the estuary, and 14 days is also

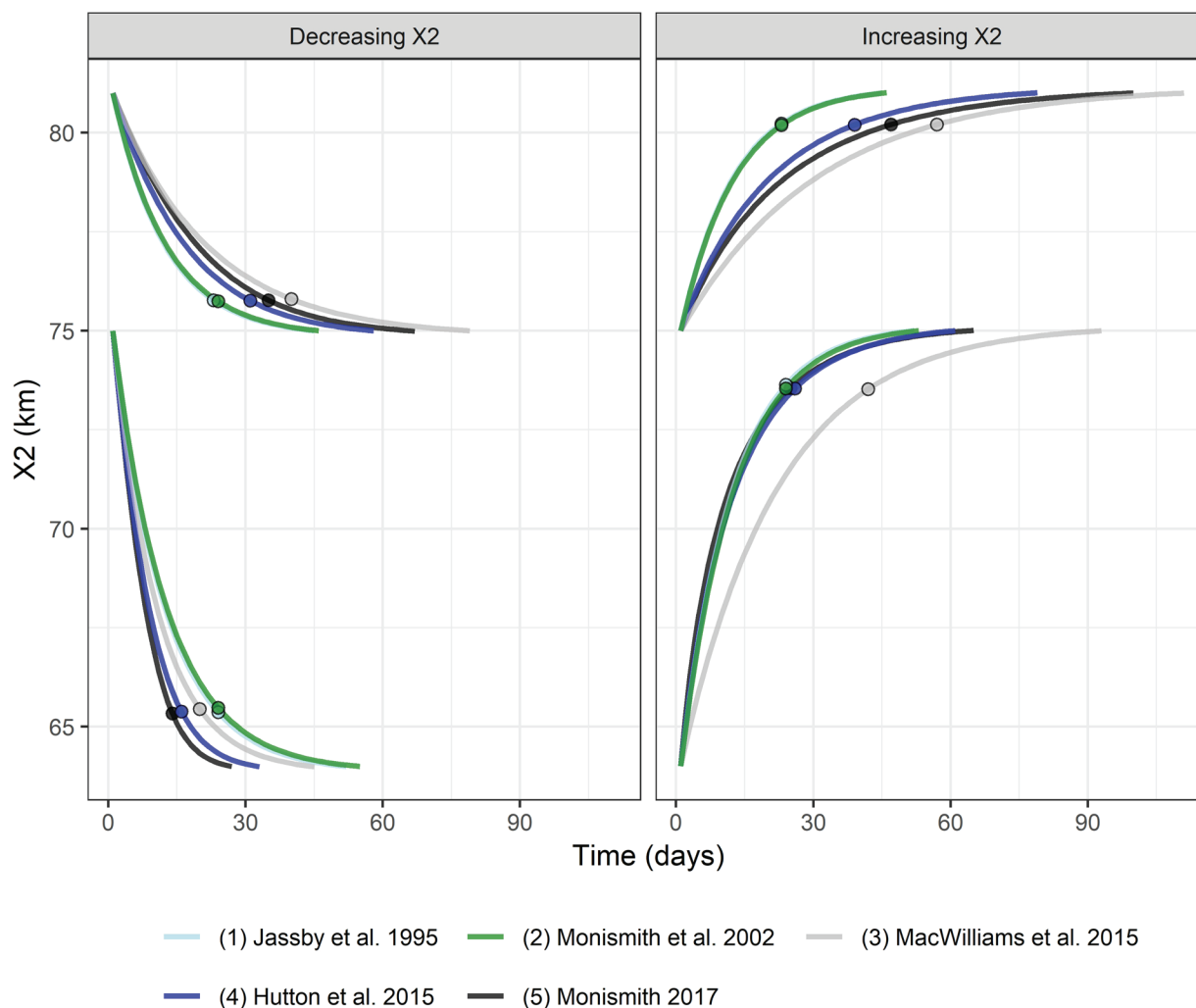


Figure 9 This figure shows the dynamic response to a multi-day step change in Delta outflow associated with each recalibrated empirical X2 model (excluding Model 6). Specifically, the figure shows the predicted response time of each model when transitioning between two steady-state positions. This test compares all possible combinations of steady transition between isohaline positions of 64, 75, or 81 km currently used for X2 regulations. The double e-folding time (86%) is used as the standard of completion, as indicated with a *dot*.

close to the characteristic time-scale of the response of the estuary. Therefore, empirical X2 model behavior at this frequency is an important characteristic, even without considering other possible characteristics at this time-scale, such as modulation of dispersion. We tested the sensitivity of the models (excluding Model 6) by imposing an outflow oscillation around three equilibrium conditions: X2=64 km, 75 km, and 81 km. A 14-day flow oscillation period was simulated with the series $A \sin\left(\frac{2\pi}{14}t\right)$, where A is 50% of the mean of the equilibrium flow values for each model, and t is time in days. We

performed this test focusing on two questions: (1) How do model responses differ at this frequency? and (2) Are model responses sufficiently nonlinear to affect mean X2 predictions? We acknowledge that the oscillation magnitude of 50% is a rather large perturbation; significant variations of this magnitude are uncommon, and the analysis was intended to elicit nonlinearity.

The responses to these oscillations differ greatly across models and flow regimes (see Figure 10). Models 1 and 2 exhibit a larger response to the flow perturbation and a positive bias after several

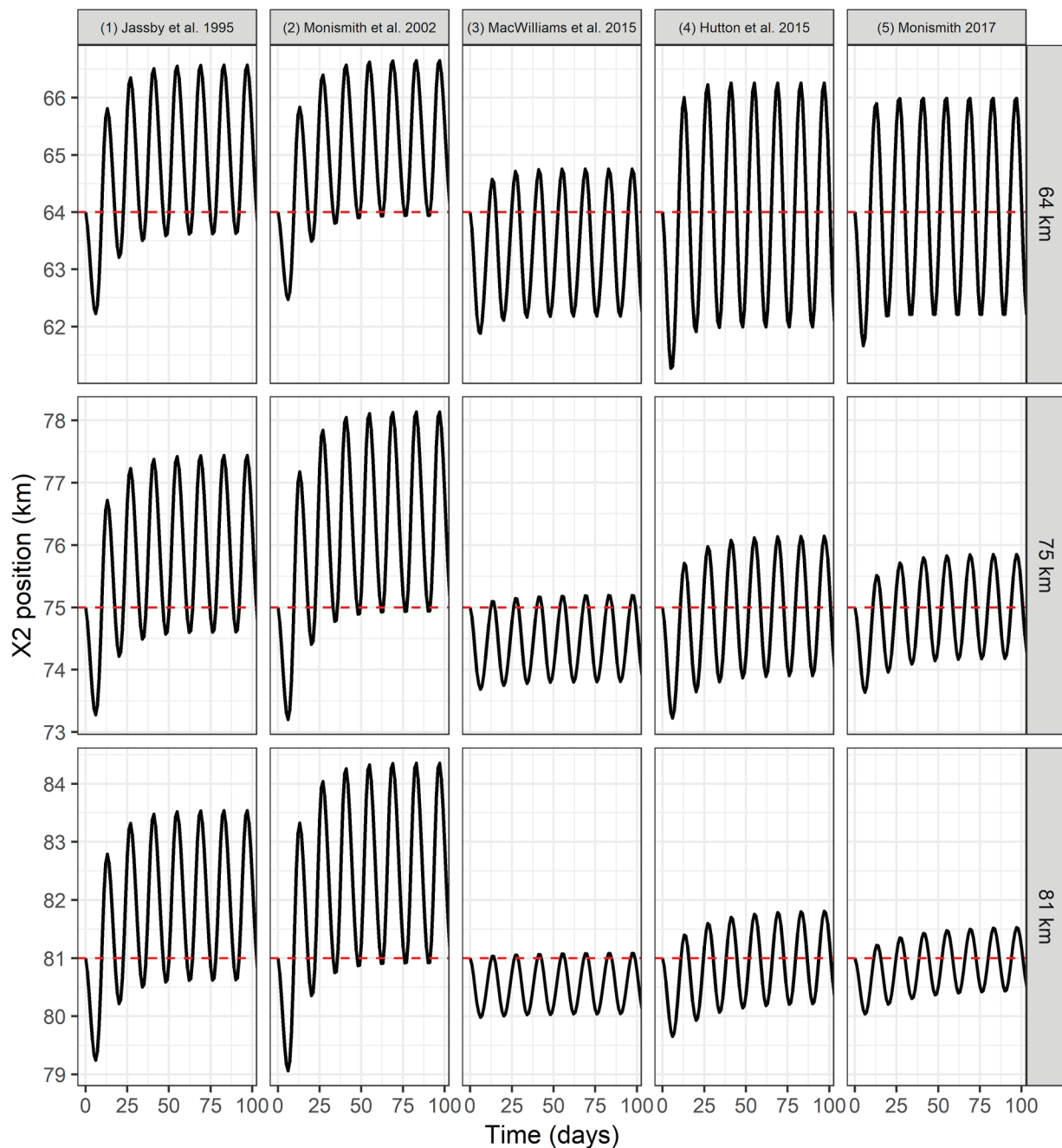


Figure 10 This figure shows the dynamic model response to 14-day period outflow oscillations around three equilibrium conditions ($X_2 = 64$ km, 75 km, and 81 km) for each recalibrated empirical X_2 model (excluding Model 6). The *dashed red lines* show the initial equilibrium X_2 position before oscillatory flow forcing begins.

cycles. Models 3, 4, and 5 exhibit similar or greater-amplitude response at the low equilibrium condition (i.e., $X_2 = 64$ km); however, this response diminishes at higher-equilibrium X_2 values. In the case of Model 5, this tendency can be explained for small amplitudes by linearizing the

model around a nominal equilibrium X_2 value and examining the time constant of the resulting first-order system. This time constant can be shown to be proportional to and crosses over 14 days as X_2 ranges between 64 km and 75 km. A longer time constant is associated with lower

frequency response to this (or any) frequency. Arguments by Chen (2015) and Monismith (2017), as well as the present experiments, agree that these characteristics are reasonable: oscillations at this frequency are admitted when X2 is low and suppressed when it is high. Models 1 and 2 exhibit responses that stay the same or grow with X2. For large-amplitude oscillations, the asymmetry between freshening and intrusion would augment this response, which was noted by Monismith (2017), but the effect seems to be modest here.

The other striking outcome of the 14-day oscillations is the generation of mean bias (positive for Models 1 and 2, negative for Model 3). Since this is a transfer of energy from one frequency (14-day period) to the mean (zero frequency), it is of nonlinear origin. The phenomenon is easiest to explain using Model 1 as an example. As stated before, this model is of the Hammerstein class of “block stationary” models, meaning that it can be written as the combination of a static nonlinearity (log) function on the inputs, followed by an otherwise linear model for the dynamics. Because the linear dynamic system is incapable of transferring energy from one band to another, it follows that the static nonlinear flow transformation, $\log\left(Q_{\text{ref}} + A \sin\left(\frac{2\pi}{14t}\right)\right)$ in the case of

Model 1, is responsible for altering the mean. To the best of our knowledge, this side effect of log transformation has not been investigated or shown to be desirable.

Besides the oscillations we investigated in this experiment, smaller oscillations at fortnightly scales are apparent in Figure 3 under higher salinity conditions. These are produced by rectified tidal processes and cannot be reproduced with NDOI inputs, which are inherently non-tidal. Model 6, which requires tidal range as well as outflow as model input, is better able to transition from one cause of fortnightly wiggles to the other.

Outflow Perturbation

Our final sensitivity analysis explored a proposition by Monismith (2017) that uniformly perturbed outflows might yield better empirical

performance. Figure 11 presents results from our evaluation of empirical X2 model performance when subjected to outflow time-series that were adjusted upward or downward by 2,000 cfs and then recalibrated. For each model (except Model 6), residual density histograms are presented by month for each outflow scenario. The range of the x-axis on each panel is restricted to +/- 10 km for visual clarity, even though a small number of points have larger absolute residuals that fall outside that range.

Model 5 exhibits the greatest difference in residual distribution between flow scenarios. Model 5 generally fits better to the reduced outflow scenario than to the base case, which may have to do with its tendency to over-respond to freshets. The remaining models show few systematic differences in residual distribution between outflow scenarios. Finally, we note that the models show common patterns of systematic seasonal bias; this observation is explored in a companion paper (Hutton et al. 2021, this volume).

DISCUSSION

Several empirical X2 models for the estuary, conceived as improvements over the seminal work published in Jassby et al. (1995) (Model 1), have been published over the past 2 decades. Here, we surveyed and compared the performance of Model 1 with other empirical models published in Monismith et al. (2002) (Model 2), MacWilliams et al. (2015) (Model 3), Hutton et al. (2015) (Model 4), and Monismith (2017) (Model 5). As part of this work, we recalibrated the models to a common data set and tested them for accuracy, mean, and transient behavior. A sixth model (Rath et al. 2017), developed with a more complex machine learning formulation, was included in this survey to provide an empirical point of comparison with the other five empirical models on skill but was not subjected to the full suite of model tests.

We found the empirical models to perform reasonably well across a wide range of conditions that were observed over the last 5 decades (WYs 1968 through 2017), a period that followed completion of large water management facilities

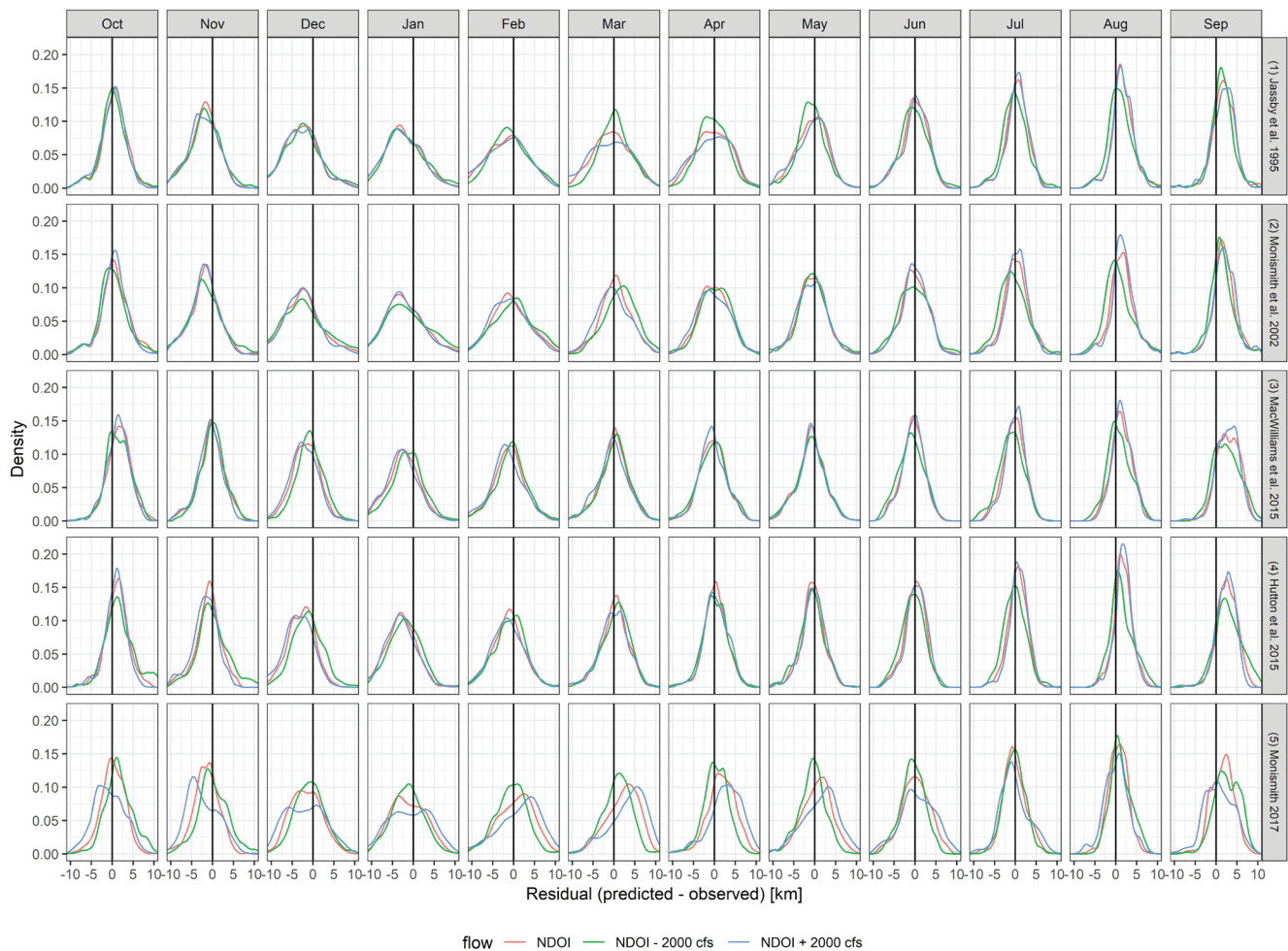


Figure 11 This figure shows posterior residuals for each empirical X2 model (excluding Model 6) when subjected to outflow time-series that were uniformly adjusted upward or downward by 2,000 cfs and further recalibration. The plot shows residual density estimates by month for three outflow scenarios: baseline NDOI (red), decreased NDOI (green), and increased NDOI (blue). The range of the x-axis on each panel is restricted to +/- 10 km for visual clarity, even though a small number of points have larger absolute residuals that fall outside that range.

in the estuary and its watershed. Relative skill varied with hydrologic conditions, but across the span of years, seasons, and outflow conditions, it is not evident that a single empirical model is superior to all others. Model 6 showed the best overall empirical performance with somewhat lower residuals; however, it has a more complex formulation with additional inputs.

As an ensemble, the models agree on the basic stability of the estuary’s outflow-salinity relationship over the last century. Although this work generally focused on the last 50 years, we conducted limited model testing using an X2

data set that extends back to WY 1922 (Hutton et al. 2015). This century-long period has seen extensive change in virtually all aspects of the study area’s land and water use as well as the estuary’s bathymetry. Despite these changes, the empirical X2 models surveyed in this work generally perform robustly over this period, with no obvious meaningful patterns in the model residuals. The exception to this finding is that the formulations of Models 1, 2, and 3 preclude their use in effectively simulating extremely low outflow conditions observed in the early part of the data record. This finding suggests that the outflow-salinity relationship in the estuary

has evolved very modestly. As with the 50-year evaluation (WYs 1968 through 2017), no single model appears to be uniformly superior when performance is evaluated by month. Overall, the empirical models are shown to be adequate for the long-term analysis of different hydrologic conditions (noting the exception of extremely low outflows), especially when looking at the past, where sea level change was limited (1.8 mm/year over the 20th century; Ryan and Noble 2007). Indeed, in an analysis of X2 trends with a constant sea level (corresponding to a 1920 level of development), Rath et al. (2017) estimated that the previous century's sea level rise affected X2 by 1 to 2 km or less under most flow conditions, and even this sensitivity may have been confounded with other time-varying inputs since the main trend of sea level is monotonic over time.

In contrast to the finding of outflow-salinity stability over the last century (i.e., during a period of continued development), Andrews et al. (2017) found through 3-D hydrodynamic modeling that salt intrusion in the pre-development estuary (circa 1850) was slightly more sensitive to outflow and responded faster to changes in outflow than in the contemporary system. They reported that tidal trapping and other unsteady processes were more important in the pre-development estuary than in the contemporary one. As part of their analysis, Andrews et al. (2017) re-calibrated the empirical X2 relationship of Hutton et al. (2015) to represent daily-averaged bottom salinity for both pre-development and contemporary simulations. Gross et al. (2018), in their comparison of pre-development, pre-project (circa 1920), and contemporary outflow and salinity scenarios, built upon the Andrews et al. (2017) work by extending the empirical simulation period from 3 years to 82 years. Confirming our findings for outflow-salinity stability over the last century, Gross et al. (2018) showed that the empirical X2 relationship developed by Andrews et al. (2017) for contemporary conditions provided a good fit to observed pre-project data that spanned WYs 1922 through 1941.

We found that the empirical X2 models exhibit different dynamic responses to outflow changes;

these responses essentially fall into two groups. Models 3, 4, and 5 demonstrate a slower response under more saline conditions, both to step changes and to fluctuations. The time constant of those models is considerably longer than 14 days when X2 is upstream of 75 km, and is closer to 14 days when X2 is downstream of 64 km. These models not only respond faster under less saline conditions; they accelerate the process of freshening. In contrast, the nonlinear responses in Models 1 and 2 follow logarithmic or power-law relationships inherent in their formulations. While we believe that the empirical skill of Models 1 and 2 are adequate, their application should be limited to situations where these dynamic and mean behaviors are acceptable.

An alternative analytical approach for evaluating model uncertainty, not covered in this work, would be to focus on a single empirical model and consider the propagation of uncertainty in parameters—including errors and biases in the measurements used for calibration—to develop ranges of X2 predictions. Such an approach was beyond our scope here but may be studied in future work to better understand flow-X2 behavior in the estuary.

As part of this work's sensitivity analyses, we evaluated the consequences of calibrating these empirical X2 models to an uncertain outflow history. We found that four of the five models (excepting Model 5) could be calibrated equally well to a uniformly perturbed outflow time-series. Recalibration produced new fitting constants for each model, with any potential errors or improvements in the outflow time-series simply absorbed in the parameterization. A corollary of this finding is that the inverse capabilities of the models are limited. The possibility of using salinity to infer Delta outflow estimates has been proposed (CDWR 2016; Fleenor et al. 2016), and Monismith (2017) essentially took this step by altering the flow record. The models do not seem to warrant faith that they can be used to identify structural flaws in outflow over the long run, but they agree very well when applied to infer seasonal anomalies or departures—a point taken

up in a companion paper (Hutton et al. 2021, this volume).

The above discussion, including evaluation of long hydrologic records and inverse applications using empirical X2 models, illustrates the possible utility of an ensemble approach. While it is naïve to ignore the fact that most users of this class of model desire an analysis tool that is quick and easy to use, incorporating more than one of the above models into an analysis framework can be accomplished with little added difficulty. The models are generally easy to implement in a spreadsheet, and together they provide a range of X2 estimates with some diversity of approach and across calibration niches. Of course, model diversity is limited because they share many similarities in their assumptions: single dominant power law or scaling with flow, homogenous spatial treatment with no consideration of bathymetry, omission of flow control structures (e.g., Suisun Marsh Salinity Control Structure and Delta Cross Channel), no treatment of wind and—with the exception of Model 6—no modulation of mean processes by the tide. Furthermore, given the difficulties inherent in defining X2 or accurately measuring Delta outflow, the resulting uncertainty in these estimates may be a bigger limitation on accurately predicting the position of the X2 isohaline than the choice of empirical model. Despite these limitations, the models perform well individually and as a group, and are expected to remain in application for years to come.

ACKNOWLEDGEMENTS

The authors wish to thank Drs. Michael MacWilliams and Ed Gross for their valuable comments and suggestions, which improved the quality of this manuscript. The Metropolitan Water District of Southern California provided funding for JSR, PHH, and SBR. The California Department of Water Resources provided funding for ESA.

REFERENCES

- Alber M. 2002. A conceptual model of estuarine freshwater inflow management. *Estuaries Coasts*. [accessed 2021 Sep 15];25(6):1246-1261. <https://doi.org/10.1007/BF02692222>
- Andrews S, Gross E, Hutton PH. 2017. Modeling salt intrusion in the San Francisco Estuary prior to anthropogenic influence. *Cont Shelf Res*. [accessed 2021 Sep 15];146:58-81. <https://doi.org/10.1016/j.csr.2017.07.010>
- Ateljevich E, Nam K, Zhang Y, Wang RF, Shu Q. 2014. Bay-Delta SELFE calibration overview, methodology for flow and salinity estimates in the Sacramento-San Joaquin Delta and Suisun Marsh. Chapter 7 in: 35th Annual Progress Report to the State Water Resources Control Board, June. [accessed 2021 Sep 15]. Sacramento (CA): California Department of Water Resources. Available from: <https://data.ca.gov/dataset/methodology-for-flow-and-salinity-estimates-in-the-sacramento-san-joaquin-delta-and-suisun-mars/resource/36ade542-2c2a-421c-ad87-7bcf581de909>
- Bernstein BB. 2012. Workshop summary: technical workshop on estuarine habitat in the Bay Delta Estuary. Appendix 4. March 27th 2012. [unknown]: US Environmental Protection Agency. [accessed 2021 Sep 15]. Available from: https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/docs/cmnt081712/dfg/cdfgbernstein2012.pdf
- Billings SA. 2013. Nonlinear system identification: NARMAX Methods in the time, frequency, and spatio-temporal domains. Chichester, West Sussex (UK): Wiley. 555 p.
- [CDPW] California Department of Public Works. 1931. Variation and control of salinity in Sacramento-San Joaquin Delta and upper San Francisco Bay. Bulletin 27. [accessed 2021 Sep 15]. Available from: <http://www.water.ca.gov/waterdatalibrary/docs/historic/bulletins.cfm>
- [CDWR] California Department of Water Resources. 2016. On estimating Net Delta Outflow (NDO): approaches to estimating NDO in the Sacramento-San Joaquin Delta. Sacramento (CA): California Department of Water Resources, Bay-Delta Office. [accessed 2021 Sep 15]. Available from: https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/deltaflow/docs/ndo_report_march2016.pdf

- [CDWR]. 2020. DSM2: Delta Simulation Model II. [accessed 2021 Sep 15]. Sacramento (CA): California Department of Water Resources. Available from: <https://water.ca.gov/Library/Modeling-and-Analysis/Bay-Delta-Region-models-and-tools/Delta-Simulation-Model-II>
- [CSWRCB] California State Water Resources Control Board. 2000. Water Rights Decision 1641. [accessed 2021 Sep 15]. Sacramento (CA): California State Water Resources Control Board. Available from: https://www.waterboards.ca.gov/waterrights/board_decisions/adopted_orders/decisions/d1600_d1649/wrd1641_1999dec29.pdf
- Chen SN. 2015. Asymmetric estuarine responses to changes in river forcing: a consequence of nonlinear salt flux. *J Phys Oceanogr.* [accessed 2021 Sep 15];45:2836-2847. <https://doi.org/10.1175/JPO-D-15-0085.1>
- Cheng RT, Casulli V, Gartner JW, 1993. Tidal, Residual, Intertidal Mudflat (TRIM) Model and its application to San Francisco Bay, California. *Estuar Coast Shelf Sci.* [accessed 2021 Sep 15];36(3):235-280. <https://doi.org/10.1006/ecss.1993.1016>
- Chua VP, Fringer OB. 2011. Sensitivity analysis of three-dimensional salinity simulations in North San Francisco Bay using the unstructured-grid SUNTANS model. *Ocean Model.* [accessed 2021 Sep 15];39(3-4):332-350. <https://doi.org/10.1016/j.ocemod.2011.05.007>
- Cloern JE, Jassby AD, Schraga TS, Nejad E, Martin C. 2017. Ecosystem variability along the estuarine salinity gradient: examples from long-term study of San Francisco Bay. *Limnol Oceanogr.* [accessed 2021 Sep 15];62(S1). <https://doi.org/10.1002/lno.10537>
- Denton RA. 1993. Accounting for antecedent conditions in seawater intrusion modeling—applications for the San Francisco Bay-Delta. Proceedings of the 1993 Hydraulic Division National Conference, American Society of Civil Engineers; July 25–30, San Francisco. [accessed 2021 Nov 15]. Available from: https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/deltaflow/docs/exhibits/ccwd/spprt_docs/ccwd_denton_1993.pdf
- Fernández-Delgado C, Baldó F, Vilas C, García-González D, Cuesta JA, González-Ortegón E, Drake P, 2007. Effects of the river discharge management on the nursery function of the Guadalquivir River estuary (SW Spain). *Hydrobiologia.* [accessed 2021 Sep 15];587(1):125-136. <https://doi.org/10.1007/s10750-007-0691-9>
- Feyrer F, Nobriga ML, Sommer TR. 2007. Multidecadal trends for three declining fish species: habitat patterns and mechanisms in the San Francisco Estuary, California. USA. *Can J Fish Aquat.* [accessed 2021 Sep 15];64(4):723-734. <https://doi.org/10.1139/F07-048>
- Fleener W, Goodwin P, Monsen N, Ruhl, C. 2016. An independent peer review report for the Delta Science Program on the DWR report on estimating Net Delta Outflow (NDO): approaches to estimating NDO in the Sacramento-San Joaquin Delta. Sacramento (CA); Delta Stewardship Council. [accessed 2017 Aug 20]. Available from: <https://deltacouncil.ca.gov/delta-science-program/scientific-peer-review>
- Gross ES, Hutton PH, Draper AJ. 2018. A comparison of outflow and salt intrusion in the pre-development and contemporary San Francisco Estuary. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];16(3). <https://doi.org/10.15447/sfews.2018v16iss3art6>
- Gross ES, Koseff JR, Monismith SG. 1999. Three-dimensional salinity simulations of south San Francisco Bay. *J Hydraul Eng.* [accessed 2021 Sep 15];125(11). [https://doi.org/10.1061/\(ASCE\)0733-9429\(1999\)125:11\(1199\)](https://doi.org/10.1061/(ASCE)0733-9429(1999)125:11(1199))
- Gross ES, MacWilliams ML, Kimmerer WJ. 2009. Three-dimensional modeling of tidal hydrodynamics in the San Francisco Estuary. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];7(2). <https://doi.org/10.15447/sfews.2009v7iss2art2>
- Harder JA. 1977. Predicting estuarine salinity from river inflows. *J Hydraul Div.* [accessed 2021 Sep 15];103(8). <https://doi.org/10.1061/JYCEAJ.0004807>

- Hutton PH. 2014. Delta salinity gradient (DSG) model, version 1.0 model documentation. Sacramento (CA): Metropolitan Water District [accessed 2021 Sep 15]. Available from: https://www.baydeltalive.com/assets/588ee18bdb51ef1619ac6fd28b97f694/application/pdf/Hutton_2014_DSG_Report.pdf
- Hutton PH, Rath J, Chen L, Unga M, Roy SB. 2015. Nine decades of salinity observations in the San Francisco Bay and Delta: modeling and trend evaluation. *J Water Resour Plan Manage.* [accessed 2021 Sep 15];142(3). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000617](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000617)
- Hutton PH, Rath JS, Ateljevich E, Roy SB. 2021. Apparent seasonal bias in Delta outflow estimates as revealed in the historical record of the San Francisco Estuary: implications for Delta Net Channel depletion estimates. *San Franc Estuary Watershed Sci.* [accessed 2021 Dec 7];19(4). <https://doi.org/10.15447/sfews.2021v19iss4art4>
- Jassby AD, Kimmerer WJ, Monismith SG, Armor C, Cloern JE, Powell TM, Schubel JR, Vendliniski TJ. 1995. Isohaline position as a habitat indicator for estuarine populations. *Ecol Appl.* [accessed 2021 Sep 15];5(1):272-289. <https://doi.org/10.2307/1942069>
- Kimmerer WJ. 2004. Open water processes of the San Francisco Estuary: from physical forcing to biological responses. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 5];2(1). <http://escholarship.org/uc/item/9bp499mv>
- Kimmerer WJ, Gross ES, MacWilliams ML. 2009. Is the response of estuarine nekton to freshwater flow in the San Francisco Estuary explained by variation in habitat volume? *Estuaries Coast.* [accessed 2021 Sep 15];32(2):375-389. <https://doi.org/10.1007/s12237-008-9124-x>
- Kimmerer WJ, MacWilliams ML, Gross ES. 2013. Variation of fish habitat and extent of the low salinity zone with freshwater flow in the San Francisco Estuary. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];11(4). <https://doi.org/10.15447/sfews.2013v11iss4art1>
- Lerczak JA, Geyer WR, Ralston DK. 2009. The temporal response of the length of a partially stratified estuary to changes in river flow and tidal amplitude. *J Phys Oceanogr.* [accessed 2021 Sep 15];39(4):915-933. <https://doi.org/10.1175/2008JPO3933.1>
- Lund J, Medellín-Azuara J, Durand J, Stone K. 2018. Lessons from California's 2012–2016 drought. *J Water Resour Plan Manage.* [accessed 2021 Sep 15];144(10). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000984](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000984)
- Lund JR, Hanak E, Fleenor W, Bennett WA, Howitt RE, Mount JF, Moyle PB. 2010. Comparing futures for the Sacramento-San Joaquin Delta. Berkeley (CA): University of California Press. 231 p.
- Luoma SN, Dahm CN, Healey M, Moore JN. 2015. Challenges facing the Sacramento–San Joaquin Delta: complex, chaotic, or simply cantankerous?. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];13(3). <https://doi.org/10.15447/sfews.2015v13iss3art7>
- MacCready P. 1999. Estuarine adjustment to changes in river flow and tidal mixing. *J Phys Oceanogr.* [accessed 2021 Sep 15];29:708–726. [https://doi.org/10.1175/1520-0485\(1999\)029%3C0708:EATCIR%3E2.0.CO;2](https://doi.org/10.1175/1520-0485(1999)029%3C0708:EATCIR%3E2.0.CO;2)
- Mac Nally R, Thomson JR, Kimmerer WJ, Feyrer F, Newman KB, Sih A, Bennett WA, Brown L, Fleishman E, Culberson SD, et al. 2010. Analysis of pelagic species decline in the upper San Francisco Estuary using Multivariate Autoregressive Modeling (MAR). *Ecol Appl.* [accessed 2021 Sep 15];20(5):1417-1430. <https://doi.org/10.1890/09-1724.1>
- MacWilliams ML, Ateljevich ES, Monismith SG, Enright C. 2016. An overview of multi-dimensional models of the Sacramento–San Joaquin Delta. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];14(4). <https://doi.org/10.15447/sfews.2016v14iss4art2>
- MacWilliams ML, Bever AJ, Gross ES, Ketefian GS, Kimmerer WJ. 2015. Three-dimensional modeling of hydrodynamics and salinity in the San Francisco Estuary: an evaluation of model accuracy, X2, and the low-salinity zone. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];13(1). <https://doi.org/10.15447/sfews.2015v13iss1art2>

- Martyr-Koller RC, Kernkamp HWJ, van Dam A, van der Wegen M, Lucas LV, Knowles N, Jaffe B, Fregoso TA. 2017. Application of an unstructured 3D finite volume numerical model to flows and salinity dynamics in the San Francisco Bay-Delta. *Estuar Coast Shelf Sci.* [accessed 2021 Sep 15];192:86-107.
<https://doi.org/10.1016/j.ecss.2017.04.024>
- Monismith SG. 2016. A note on Delta outflow. *San Franc Estuary Watershed Sci.* [accessed 2021 Sep 15];14(3).
<https://doi.org/10.15447/sfews.2016v14iss3art3>
- Monismith SG. 2017. An integral model of unsteady salinity intrusion in estuaries. *J Hydraul Res.* [accessed 2021 Sep 15];55(8):1-17.
<https://doi.org/10.1080/00221686.2016.1274682>
- Monismith SG, Kimmerer W, Burau JR, Stacey MT. 2002. Structure and flow-induced variability of the subtidal salinity field in northern San Francisco Bay. *J Phys Oceanogr.* [accessed 2021 Sep 15];32(11).
[https://doi.org/10.1175/1520-0485\(2002\)032%3C3003:SAFIVO%3E2.0.CO;2](https://doi.org/10.1175/1520-0485(2002)032%3C3003:SAFIVO%3E2.0.CO;2)
- Murphy DD, Weiland PS. 2019. The low-salinity zone in the San Francisco Estuary as a proxy for Delta Smelt habitat: a case study in the misuse of surrogates in conservation planning. *Ecol Indic.* [accessed 2021 Sep 15];105:29-35.
<https://doi.org/10.1016/j.ecolind.2019.05.053>
- Murray Darling Basin Ministerial Council. 1999. The salinity audit of the Murray Darling Basin: a 100-year perspective. [accessed 2021 Sep 15]. Available from: http://www.mdba.gov.au/sites/default/files/archived/mdbc-salinity-reports/439_Salinity_audit_of_MDB_100_year_perspective.pdf
- Rath JS, Hutton PH, Chen L, Roy SB. 2017. A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. *Environ Model Softw.* [accessed 2021 Sep 15];93:193-208.
<https://doi.org/10.1016/j.envsoft.2017.03.022>
- Reinert TR, Peterson JT. 2008. Modeling the effects of potential salinity shifts on the recovery of striped bass in the Savannah River Estuary, Georgia-South Carolina, United States. *Environ Manage.* [accessed 2021 Sep 15];41(5):753-765.
<https://doi.org/10.1007/s00267-008-9082-x>
- Roy SB, Rath JS, Chen L, Uings MJ, Guerrero M. 2014. Salinity trends in Suisun Bay and the western Delta: October 1921-September 2012. Prepared for San Luis and Delta Mendota Water Authority and State Water Contractors. [accessed 2021 Sep 15]. Available from: <https://core.ac.uk/reader/78049090>
- Ryan HF, Noble MA. 2007. Sea level fluctuations in central California at subtidal to decadal and longer time scales with implications for San Francisco Bay, California. *Estuar Coast Shelf Sci.* [accessed 2021 Sep 15];73(3-4):538-550.
<https://doi.org/10.1016/j.ecss.2007.02.009>
- Schubel JR. 1993. Managing freshwater discharge to the San Francisco Bay-Delta Estuary: the scientific basis for an estuarine standard. San Francisco (CA): San Francisco Estuary Project. [accessed 2021 Sep 15]. Available from: https://www.epa.gov/sites/default/files/documents/sfep_1993_managing_fw_discharge_sf_bay_delta_estuary_0.pdf
- Sen PK. 1968. Estimates of the regression coefficient based on Kendall's tau. *J Amer Stat Assoc.* [accessed 2021 Sep 15];63(224):1379-1389.
<https://doi.org/10.1080/01621459.1968.10480934>
- Sklar FH, Browder JA. 1998. Coastal environmental impacts brought about by alterations to freshwater flow in the Gulf of Mexico. *Environ Manage.* [accessed 2021 Sep 15];22(4):547-562.
<https://doi.org/10.1007/s002679900127>
- Tamburello N, Connors BM, Fullerton D, Phillis CC. 2019. Durability of environment-recruitment relationships in aquatic ecosystems: insights from long-term monitoring in a highly modified estuary and implications for management. *Limnol Oceanogr.* [accessed 2021 Sep 15];64(2):1-17.
<https://doi.org/10.1002/lno.11037>