UC Davis
San Francisco Estuary and Watershed Science

Title
Three-Dimensional Modeling of Hydrodynamics and Salinity in the San Francisco Estuary: An Evaluation of Model Accuracy, X2, and the Low–Salinity Zone

Permalink
https://escholarship.org/uc/item/7x65r0tf

Journal
San Francisco Estuary and Watershed Science, 13(1)

Authors
MacWilliams, Michael L.
Bever, Aaron J.
Gross, Edward S.
et al.

Publication Date
2015

DOI
https://doi.org/10.15447/sfews.2015v13iss1art2

Supplemental Material
https://escholarship.org/uc/item/7x65r0tf#supplemental

Peer reviewed

Three-Dimensional Modeling of Hydrodynamics and Salinity in the San Francisco Estuary: An Evaluation of Model Accuracy, X2, and the Low-Salinity Zone

Michael L. MacWilliams*[1], Aaron J. Bever[1], Edward S. Gross[2], Gerard S. Ketefian[2], and Wim J. Kimmerer[3]

ABSTRACT

The three-dimensional UnTRIM San Francisco Bay–Delta model was applied to simulate tidal hydrodynamics and salinity in the San Francisco Estuary (estuary) using an unstructured grid. We compared model predictions to observations of water level, tidal flow, current speed, and salinity collected at 137 locations throughout the estuary. A quantitative approach based on multiple model assessment metrics was used to evaluate the model's accuracy for each comparison. These comparisons demonstrate that the model accurately predicted water level, tidal flow, and salinity during a 3-year simulation period that spanned a large range of flow and salinity conditions. The model is therefore suitable for detailed investigation of circulation patterns and salinity distributions in the estuary.

The model was used to investigate the location, and spatial and temporal extent of the low-salinity zone (LSZ), defined by salinity between 0.5 and 6 psu. We calculated X2, the distance up the axis of the estuary to the daily-averaged 2-psu near-bed salinity, and the spatial extent of the LSZ for each day during the 3-year simulation. The location, area, volume,

and average depth of the low-salinity zone varied with X2; however this variation was not monotonic and was largely controlled by the geometry of the estuary.

We used predicted daily X2 values and the corresponding daily Delta outflow for each day during the 3-year simulation to develop a new equation to relate X2 to Delta outflow. This equation provides a conceptual improvement over previous equations by allowing the time constant for daily changes in X2 to vary with flow conditions. This improvement resulted in a smaller average error in X2 prediction than previous equations. These analyses demonstrate that a well-calibrated three-dimensional (3-D) hydrodynamic model is a valuable tool for investigating the salinity distributions in the estuary, and their influence on the distribution and abundance of physical habitat.

KEYWORDS

San Francisco Bay, Sacramento–San Joaquin Delta, hydrodynamic modeling, UnTRIM, calibration, model accuracy, model skill, low-salinity zone (LSZ), X2

INTRODUCTION

The ecology of the San Francisco Estuary (estuary) is strongly influenced by tidal hydrodynamics and the distribution of salinity on hourly to interannual time-scales. The salinity field limits the physical habitat available to several species which occupy specific salinity ranges (Feyrer et al. 2007; Nobriga et al. 2008; Kimmerer et al. 2009). Salinity stratification, which varies on spring–neap and seasonal time-scales (e.g., Monismith et al. 1996; Burau et al. 1998; Kimmerer et al. 1998), reduces vertical mixing which can profoundly affect primary productivity (Cloern 1984) and other ecological processes. The tidal and tidally-averaged circulation pattern can affect the recruitment and retention of organisms (Kimmerer

et al. 1998, 2014) and the likelihood of entrainment by export facilities that pump freshwater from the estuary.

Abundance or survival of several estuarine biological populations in the estuary have historically been positively related to freshwater flow, as indexed by the position of the daily-averaged 2-psu isohaline near the bed, or X2 (Jassby et al. 1995). Many hypotheses have been proposed that relate freshwater flow to abundance of organisms in the estuary (Kimmerer 2002). One such hypothesis is that abundance of estuarine species varies with the area or volume of low-salinity habitat, which is related to X2 (Feyrer et al. 2007; Nobriga et al. 2008; Kimmerer et al. 2009, 2013). The exploration of this hypothesis required prediction of the salinity field for a range of Delta outflows in order to calculate X2 and the area, volume, and depth of water within specific salinity ranges for multiple species (Kimmerer et al. 2009, 2013).

The estuary encompasses San Francisco Bay (Central Bay and South Bay), San Pablo Bay, Suisun Bay, and
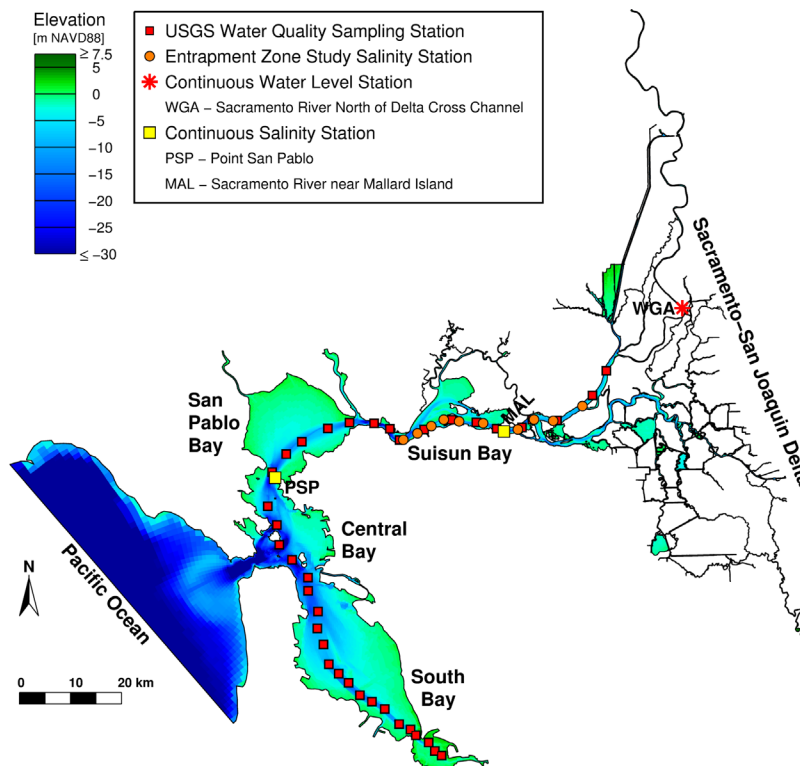

Figure 1 Locations of data collection stations in the San Francisco Estuary at which comparisons between observations and model predictions are shown in this paper. Primary sub-embayments of the San Francisco Estuary are also labeled.

the Sacramento–San Joaquin Delta (Figure 1). Here, we apply the term "SF Bay" to refer to San Francisco Bay, San Pablo Bay, Carquinez Strait, and Suisun Bay. The estuary is mesotidal with salinity conditions that typically vary from well-mixed to partially stratified (Monismith et al. 2002). Wind forcing can also exert a strong influence on observed circulation in the estuary (Cheng et al. 1997). Substantial lateral and vertical variability in velocity and salinity have been observed in field studies (Stacey et al. 2001; Fram et al. 2007).

To test hypotheses relating ecological processes to the extent of physical habitat, X2, or estuarine circulation, detailed information on hydrodynamics and salinity throughout the estuary must be established on fine spatial and temporal scales. Estuarine circulation varies with tidal forcing, wind, and freshwater inflows. This spatially and temporally heterogeneous forcing interacts with the bathymetry and salinity field of the estuary to produce complex patterns of circulation, salinity, and turbulent mixing.

The complex circulation patterns in the estuary cannot be fully described by field observations and are often three dimensional, such that a 3-D hydrodynamic model is necessary to accurately characterize and predict circulation. Further, the model must be validated using a sufficiently long simulation period to span a wide range of tidal, wind, and freshwater flow conditions to demonstrate that the model adequately represents conditions throughout the estuary.

Previous modeling studies in the estuary that used the TRIM3D model to evaluate the relationship between outflow and X2 used a Cartesian grid (Kimmerer et al. 2009; Gross et al. 2009). Although these analyses yielded valuable information, the limitations imposed by using a structured Cartesian grid to resolve the estuarine geometry in the TRIM3D model restricted the conclusions that could be drawn. In particular, the TRIM3D model could not represent most of the Delta using the 200-m grid resolution required to make simulations of the full estuary computationally feasible at the time.

In this study, we applied the unstructured UnTRIM model (Casulli and Zanolli 2002, 2005) to the estuary (Figure 2). The UnTRIM San Francisco Bay–Delta

model (UnTRIM Bay–Delta model) is a 3-D hydrodynamic model of the estuary (MacWilliams and Gross 2007; MacWilliams et al. 2008, 2009). The use of an unstructured mesh allows for gradually varying grid cell sizes, beginning with large grid cells in the Pacific Ocean and gradually transitioning to finer grid resolution in the smaller channels of the Sacramento–San Joaquin Delta. This approach offers significant advantages both in terms of numerical efficiency and accuracy relative to a structured grid approach, by allowing for local grid refinement for detailed analysis of local hydrodynamics, while still incorporating the overall hydrodynamics of the entire estuary in a single model. The model has been applied in studies in the estuary for the California Department of Water Resources (CDWR), the U.S. Bureau of Reclamation (USBR), the U.S. Geological Survey (USGS), and the U.S. Army Corps of Engineers (USACE) (e.g., MacWilliams and Cheng 2006; MacWilliams and Gross 2007, 2010; MacWilliams et al. 2008, 2009, 2012a, 2012b; Bever and MacWilliams 2013; MacWilliams and Bever 2013).

There were three primary motivations for the analysis presented here. First, we wanted to develop a quantitative approach to document the accuracy of the UnTRIM Bay–Delta model, which has been used as the basis for several recent analyses (e.g., Kimmerer et al. 2013; Kimmerer et al. 2014). We compared salinity predictions from the UnTRIM simulation to previous results from the TRIM3D model to quantify the improvement in salinity predictions possible with a more refined model grid. We used several different quantitative metrics to assess the accuracy of model predictions of water level, tidal flows and salinity. We then proposed model accuracy thresholds to facilitate comparisons between different models or among different simulation periods using the same model. Long-term changes expected in the estuary from climate change and human activities, notably changes to the physical configuration of the estuary, are likely to result in a very different salinity distribution than now exists. A highly-resolved, well-calibrated, and well-documented model is essential to confidently predict flow and salinity under future conditions.

Second, we were interested in exploring the distribution of low-salinity water in the estuary as a proxy
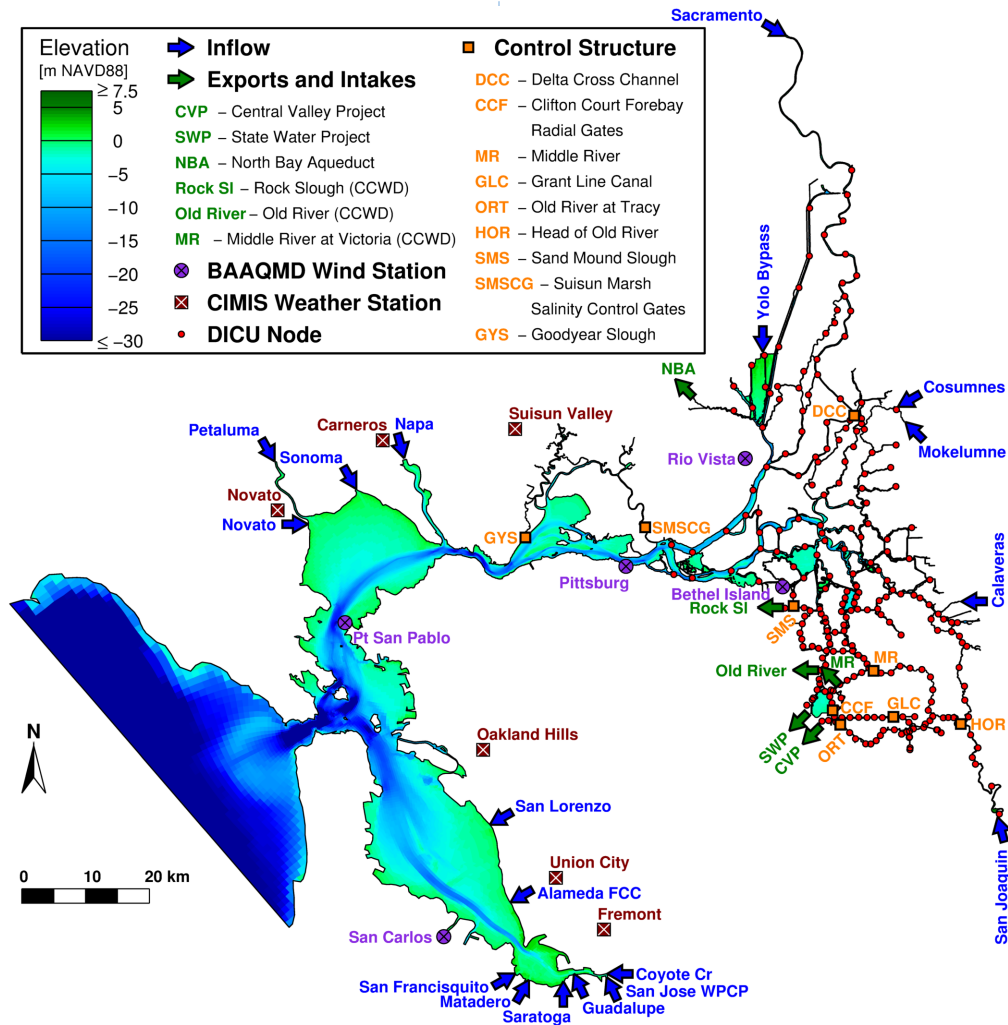
Figure 2 UnTRIM San Francisco Bay–Delta model domain, bathymetry, and locations of model boundary conditions which include inflows, water intake and export facilities, wind stations, evaporation and precipitation from CIMIS weather stations, Delta Island Consumptive Use (DICU), and flow control structures

for the physical habitat of Delta Smelt and other fishes (Feyrer et al. 2007; Kimmerer et al. 2009, 2013). The State Water Resources Control Board established standards for salinity in the estuary based on the freshwater outflows from the Delta necessary to maintain X2 at specific locations (SWRCB 2000). This regulation is based on observations that the abundance or survival of several estuarine biological populations in the estuary is positively related to freshwater flow (Jassby et al. 1995; Kimmerer et al. 2009, 2013). The Biological Opinion for Delta Smelt (*Hypomesus transpacificus*) calls for efforts to increase outflow to enlarge the area of habitat with

suitable salinity for this fish (USFWS 2008). Thus, there is considerable interest in determining the mechanisms that underlie the relationship between physical habitat and fish abundance so that the standards could be made more protective or more efficient in their use of freshwater. Therefore, we used the results of the 3-year model simulation to predict the spatial extent of low-salinity habitat and investigate how the area, depth, and volume of the low-salinity zone (LSZ) varied as a function of X2.

Third, we wanted to develop an improved equation to represent the relationship between X2 and Delta outflow. X2 has previously been determined by interpo-

lation of observed surface salinity between monitoring stations and by prediction from an autoregressive model that relates X2 to Delta outflow (Jassby et al. 1995). Although this relationship has been widely applied, the approach has several drawbacks. The equation was developed to achieve a good fit but is not based on physical theory (Monismith et al. 2002). The interpolation used data from only a handful of stations that did not include the entire range of X2, and made some assumptions about the (nonlinear) shape of the relationship of salinity to distance near freshwater. The estimation of X2 from surface salinity also required an assumption of a constant amount of salinity stratification, which introduces errors because stratification is variable. In addition, the time-series model developed by Jassby et al. (1995) uses a time constant that does not change with flow, which is contrary to understanding about the physical response of the estuary to flow. Finally, Delta outflow itself is poorly known when it is small.

In contrast, the freshwater inflows to the model are specified and are therefore known exactly, and X2 in the model can be determined to any precision directly from the modeled near-bed salinity along the axis of the estuary (Figure 3). Thus, developing a relationship from predictions of X2 from a well-calibrated hydrodynamic model offers several advantages over similar relationships developed from X2 estimates that are derived from imperfect and incomplete observational data. We applied the physically based equation of Monismith et al. (2002), modified by the use of a flow-dependent time-response term to improve the accuracy of the X2 predictions.

METHODS

Hydrodynamic Model

The governing equations, numerical discretization, and numerical properties of UnTRIM are described in Casulli (1999, 2009), Casulli and Walters (2000), and Casulli and Zanolli (2002, 2005, 2007), and are not reproduced here. UnTRIM solves the 3-D Navier–Stokes equations on an unstructured grid in the horizontal plane. The boundaries between vertical layers are at fixed elevations that can be specified non-uniformly to provide increased resolution near
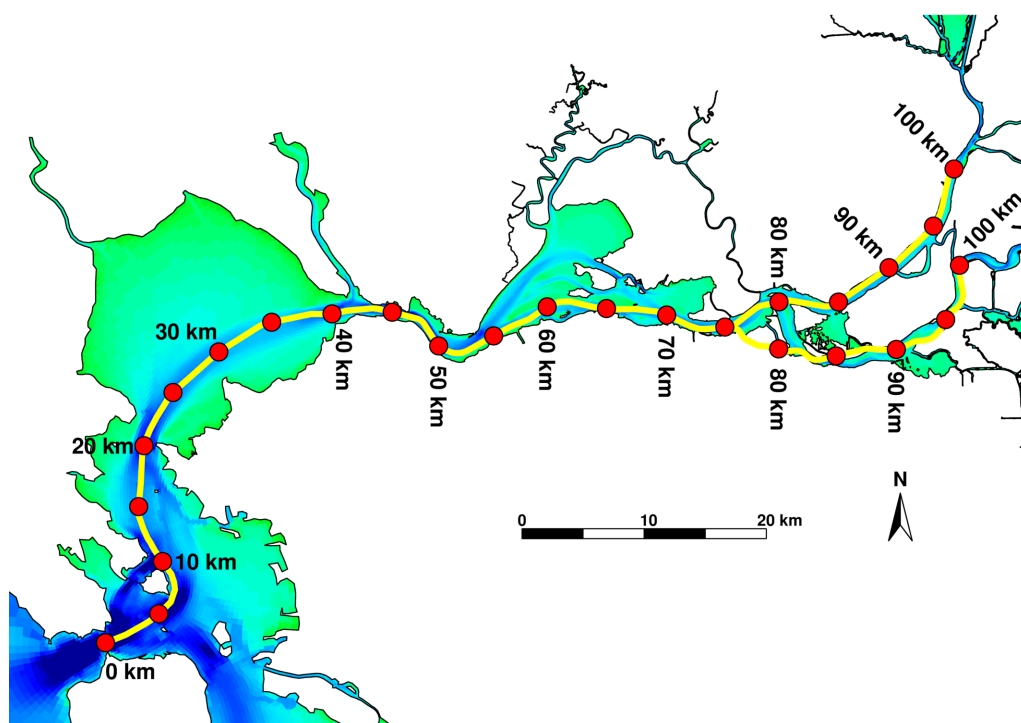

Figure 3 Transects along the axis of northern SF Bay used to calculate X2 using predicted salinity from the UnTRIM Bay–Delta model

5

the surface or at other vertical locations. Volume conservation is satisfied by a vertical integration of the incompressible continuity equation using a finite-volume discretization, and the free-surface elevation is calculated by integrating the continuity equation over the depth and using a kinematic condition at the free surface (Casulli and Cheng 1992). The gradient of surface elevation in the horizontal momentum equations and the velocity in the free-surface equation are discretized by the θ-method. When the implicitness factor, θ, is chosen in the range between 0.5 and 1.0, the stability of the resulting scheme is independent from the free-surface wave speed and bottom friction (Casulli and Cattani 1994). A value of 0.6 was used for θ in this study. The advection and horizontal viscosity terms are discretized using an Eulerian–Lagrangian approach (Casulli and Zanolli 2002). The wind stress, the vertical viscosity, and the bottom friction are discretized implicitly (Casulli and Zanolli 2002). The numerical method allows full wetting and drying of cells in the vertical and horizontal directions, and can be applied using either a hydrostatic or non-hydrostatic assumption. In this study, the hydrostatic approximation was applied. The advection-diffusion equation is solved using a finite volume scheme that guarantees both mass conservation and maximum principle, with the specification of an appropriate flux limiter function (Casulli and Zanolli 2005). The van Leer (1979) flux limiter was applied in this study. A sub-cycling approach is used to solve the advection–diffusion equation so that a further stability restriction is not imposed on the hydrodynamic model (Casulli and Zanolli 2002).

The turbulence closure model used in this study is a two-equation model consisting of a turbulent kinetic energy equation and a generic length-scale equation. The parameters of the generic length-scale (GLS) equation are chosen to yield the k-ε closure (Umlauf and Burchard 2003). The Kantha and Clayson (1994) quasi-equilibrium stability functions are used. All parameter values used in the k-ε closure are identical to those used by Warner et al. (2005b). The numerical method used to solve the equations of the turbulence closure is a semi-implicit method that results in tri-diagonal, positive–definite matrices in each water column and ensures that the turbulent variables remain positive (Deleersnijder et al. 1997).

Bathymetric and Model Grids

A high-resolution unstructured grid for the model domain was developed using the grid generator JANET (Lippert and Sellerhoff 2006). The grid was developed such that the main channels of the estuary are discretized using "orthogonal curvilinear" quadrilaterals which are aligned with the principal flow directions and along each of the navigation channels, while the remainder of the mesh is constructed using a combination of triangles and quadrilaterals (Figure 4). Within each grid cell, the cell center is defined such that the segment joining the centers of two adjacent polygons and the side shared by the two polygons have a nonempty intersection and are orthogonal to each other (see Casulli and Zanolli 2002). The grid was optimized to ensure that all grid cells are orthogonal to within 0.1 degrees. The grid resolution along the axis of the estuary varies as necessary to resolve bathymetric variability, with smaller grid cells used in narrower channels and in regions with complex bathymetry. Grid cell side lengths are approximately 1 km at the ocean boundary, 400 m at the Golden Gate, and become gradually smaller with distance in the landward direction. The horizontal grid resolution is approximately 50 to 75 m in the western Delta, and between 10 to 50 m in the central and southern Delta (Figure 4). The vertical grid resolution is 1 m to a depth of 20 m below zero NAVD88. Between 20 m below zero NAVD88 and 105 m below zero NAVD88, the vertical layer spacing gradually increases from 1 m to 5 m. The resulting grid contains 129,946 horizontal grid cells, more than 1 million 3-D grid cells, and more than 2 million active cell faces where the velocity normal to the face is computed at each time step (90 seconds). Although the model formulation does not place any stability limits on the hydrodynamic time step, a time step of 90 seconds was found to result in a good balance between the number of scalar sub-cycles within each time-step and satisfied the Courant limit based on the grid resolution and internal wave speed in stratified regions. Using this time step, the full UnTRIM Bay–Delta model simulations typically run slightly faster than 30 times real-time on a desktop workstation, which allows for the simulation of 1 calendar year in approximately 12 days.
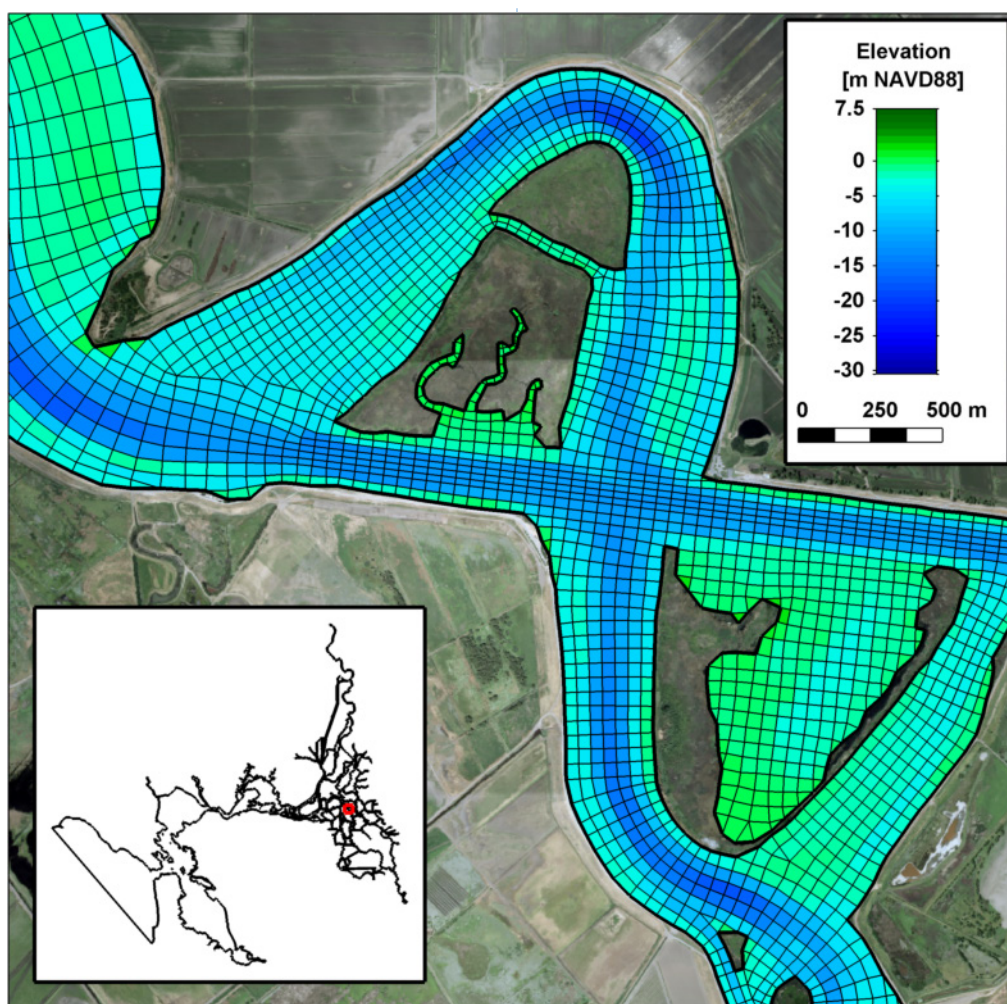
Figure 4 UnTRIM Bay-Delta model grid along the San Joaquin River between Prisoners Point and the junction with Middle River (location indicated by red square in map inset)

High-resolution bathymetric data from several sources were incorporated into the model bathymetry. A 2-m digital elevation model (DEM) based on multi-beam echo sounder data collected in 2004 and 2005 was used in the Golden Gate and adjacent portions of the coastal Pacific Ocean (CSUMB 2005). Bathymetry for the remaining portions of the coastal Pacific Ocean included in the model grid was derived from a California-vicinity bathymetric and terrestrial DEM at 90-m resolution, distributed by the California Department of Fish and Wildlife (CDFW). The primary data source for Central Bay was a 4-m DEM (USGS 2007) developed from multi-beam data collected in 1996 and 1997 (Gardner et al. 1998). High-resolution bathymetry from an acoustic hydrographic survey conducted in 2005 (Foxgrover et al. 2007) was used as the primary data source in South Bay. Bathymetry for the portions of Central Bay and South Bay not covered by the multi-beam surveys and for San Pablo Bay was derived from the 30-m DEM data generated by NOAA using NOS soundings and other bathymetry data collected from 1979 to 1985 (NOAA 1998). In Suisun Bay and the Sacramento–San Joaquin Delta, the bathymetry was developed using the USGS 10-m resolution bathymetric grid based on nearly one million depth soundings augmented by contours and aerial photography (poster presented by Smith et al. in 2003, unreferenced, see "Notes"). The USGS

bathymetry was augmented by survey data collected within the SF Bay to Stockton navigation channels and the Sacramento River Deep Water Ship Channel (Towill, Inc. 2009), and with additional bathymetric data in portions of the Delta not included in these data sets, including Liberty Island, Mildred Island, Barker Slough, and upstream portions of the Sacramento River and San Joaquin River. Each of the bathymetric data sets was projected to the Universal Transverse Mercator (UTM) North American Datum of 1983 (NAD83) coordinate system and the vertical datum was referenced to the North American Vertical Datum of 1988 (NAVD88). Although recent advances in the numerical methods used in UnTRIM allow for the specification of bathymetry at a higher resolution than the computational mesh using sub-grid bathymetry (Casulli 2009; Casulli and Stelling 2011), the current application applies the model bathymetry using the "classic" UnTRIM grid structure such that one depth value is specified on each grid face.

Simulation Period

A 3-year period from April 1, 1994 through April 1, 1997 was simulated to allow us to compare model results with extensive field data from the Entrapment Zone Study in Carquinez Strait and Suisun Bay (Burau et al. 1998; Kimmerer et al. 1998, 2002). This data set includes high-frequency velocity data from Acoustic Doppler Current Profilers (ADCPs), salinity from continuous monitoring stations, and salinity transects. The first 12 months of this period were previously used in validation of the TRIM3D model (Gross et al. 2009). Extensive validation of the UnTRIM Bay–Delta model has also been documented for several other simulation periods (MacWilliams et al. 2008, 2009; MacWilliams and Gross 2010; MacWilliams and Bever 2013).

The 3-year model simulation period spans parts of 4 water years, which encompass a large part of the historical range of Delta outflows. Water years 1994 (October 1, 1993 through September 30, 1994) through 1997 ranked 6th, 54th, 40th, and 48th, respectively, in annual mean Delta outflow over the 57-year record (CDWR 2013). January 1997

had the second-highest monthly mean outflow, and January 1, 1997 the second-highest daily outflow, in the record.

Model Boundary Conditions

Observed water levels from the NOAA San Francisco (9414290) station at the southern end of the Golden Gate Bridge were used to specify tidal water levels at the ocean boundary (Figure 2). The observations were multiplied by an amplification factor of 1.02 to account for the difference in tidal range between observed San Francisco tides and tides along the ocean boundary, and a phase lead of 47 minutes was applied to account for the phase difference between San Francisco and the model boundary. The amplification factor and phase lead were selected to minimize the phase and amplitude difference between the observed and predicted water levels at the NOAA San Francisco (9414290) station during the simulation period.

Typical observed salinity in the coastal ocean near San Francisco Bay is 33.5 psu (Dever and Lentz 1994); however, lower salinities have been observed during periods of large Delta outflow, such as those observed during winter and spring of 1995. The salinity at the ocean boundary was therefore specified using daily salinity observations from the Farallon Islands (SCCOOS 2012), approximately 20 km west of the tidal ocean boundary.

The river inflows to the model domain include tributary inflows to the Delta and to SF Bay, and discharges from water pollution control plants (Figure 2). Daily Delta inflow values were specified for six tributaries using data from the Dayflow program (CDWR 1986, 2013). Daily water export and intake flows were specified for the State Water Project (SWP), Central Valley Project (CVP), the North Bay Aqueduct (NBA), the Contra Costa Water District (CCWD) intake facilities at Rock Slough and Old River, and the Byron–Bethany Irrigation District (BBID) diversion facility (Figure 2).

Hourly evaporation and precipitation data collected by the California Irrigation Management Information

System (CIMIS) at stations west of the Delta were used to specify spatially variable evaporation and precipitation (Figure 2). Evaporation and precipitation are treated as sink and source terms, respectively, in the surface layer of each grid cell. The Delta Island Consumptive Use (DICU) model described later accounts for evaporation and precipitation in the Delta.

Wind forcing was applied at the water surface as a wind stress. The wind drag coefficient was varied based on local wind speed according to the formulation of Large and Pond (1981). Observed hourly wind speed and direction from the Bay Area Air Quality Management District (BAAQMD) from five locations (Figure 2) were used to account for spatial variability in wind velocities. At the bottom boundary, the coefficient of drag was estimated using a specified roughness coefficient (z_0) following the approach described in MacWilliams (2004). The roughness coefficient z_0 was specified according to the elevation of each grid cell edge following the approach used by Cheng et al. (1993), Gross et al. (2009) and MacWilliams and Gross (2013), and ranged from 0.001 mm to 1.0 cm, with higher roughness coefficient values specified in shallower and higher-elevation areas.

Irrigation diversions and agricultural return flows in the Delta significantly affect hydrodynamics and water quality in the Delta (CDWR 1995). The DICU model (CDWR 1995; Marvin Jung & Associates 2000) estimates channel depletions, infiltration, evaporation, precipitation, and agricultural use in the Delta. These flows are grouped into monthly estimates of net diversions, seepage, and agricultural return discharge and electrical conductivity for a total of 142 Delta sub-areas. In the 1-D DSM2 model (CDWR 2008), the DICU values for each of these sub-areas are distributed onto a total of 258 nodes on the model grid. Following the approach used by RMA (2005), each node in DSM2 was mapped to the nearest UnTRIM cell, and the corresponding DICU values for that location were applied to the UnTRIM model (Figure 2). The seepage and flow diversion components were applied as outflows from the model, while the return flow was applied at each node as an inflow with salinity corresponding to the electrical conductivity value reported in the DICU model.

Permanent control gates and temporary barriers are used in the Delta to manage water quality, protect fish migrating through the Delta, and ensure an adequate water supply for agricultural diversions in the south Delta. Nine Delta control structures are currently represented in the model (Figure 2), including the Delta Cross Channel (DCC) gates, the Clifton Court Forebay (CCF) radial gates, the Suisun Marsh Salinity Control Gates (SMSCG), the temporary barriers at Head of Old River (HOR), Old River at Tracy (ORT), Middle River (MR), and Grant Line Canal (GLC), and control structures at Sandmound Slough (SMS) and Goodyear Slough (GS) in Suisun Marsh. The model representation of each of these control structures and the seasonal timing for installation, removal, and typical operations are described in MacWilliams et al. (2009). MacWilliams and Gross (2013) describe how the operation of the CCF radial gates are implemented in the model.

The initial salinity field in San Francisco Bay was specified based on vertical salinity profiles collected by the USGS at 38 water-quality sampling stations along the axis of the estuary (Figure 1; USGS 2013) on January 18, 1994. Initial salinity conditions for the Delta were specified by interpolating linearly among available observed salinity data at the CDWR continuous salinity monitoring stations on January 18, 1994. Thus, the initial salinity field varied longitudinally and vertically, but we assumed it was laterally uniform.

The model simulation was started from quiescent conditions on January 17, 1994 to allow for hydrodynamic spin-up. The initial salinity field was specified at the beginning of the simulation, and then reset to the observed values at approximately the mid-point of the data-collection cruise on January 18, 1994, when simulated tidal currents had spun up. Sensitivity tests indicated that a 6-week spin-up period is adequate to minimize the effect of a 10% uncertainty in the initial conditions on salinity predictions in most of the estuary (MacWilliams et al. 2009); a 2.5-month spin-up period (January 17 to April 1, 1994) was simulated before the 3-year analysis period.

Evaluation of Model Accuracy

The quality of fit between the observation data and model predictions of water level, tidal flows, current speed, and salinity was assessed using a cross-correlation procedure similar to that used by RMA (2005). This approach describes the differences between two time series in terms of phase, mean, and amplitude. Statistics were calculated to quantify the differences between predicted and observed time-series data as in MacWilliams and Gross (2013). A linear regression between the time-shifted model results and observed data record was used to determine the best-fit line and coefficient of determination (r^2). The slope of the best-fit line is a measure of the ratio of tidal amplitude in modeled vs. observed data.

We used model skill and target diagrams to provide quantitative metrics for evaluating model accuracy. Willmott (1981) defined the predictive skill of a model based on the quantitative agreement between observations (O) and model predictions (M) as

$$Skill = 1 - \left[\sum_{i=1}^{N} \left| X_{Mi} - X_{Oi} \right|^2 \right] / \left[\sum_{i=1}^{N} \left(\left| X_{Mi} - \overline{X_O} \right| + \left| X_{Oi} - \overline{X_O} \right| \right)^2 \right] \tag{1}$$

where X is the variable being compared, $\overline{X}$ is the time average of X, M_i is model value at time i of N total comparison times, and O_i is the observation at time i. Perfect agreement between model results and observations yields a skill of 1. Although the Willmott (1981) model skill metric has some shortcomings (see Ralston et al. 2010), it has nevertheless been used to compare model predictions to observed data in numerous hydrodynamic modeling studies (e.g., Warner et al. 2005a; Haidvogel et al. 2008; MacWilliams and Gross 2013).

Graphical methods of evaluating model accuracy are helpful to compare: (1) the accuracy of a model with observations from a large number of stations and (2) the accuracy of different models (e.g., Bever et al. 2013). Jolliff et al. (2009) and Hofmann et al. (2011) provide a detailed description of target diagrams and their use in assessing model skill. This approach uses the bias and the unbiased Root Mean Square Difference (*ubRMSD*) to assess the accuracy of the

model predictions. The *bias* of the model estimates is calculated as

$$bias = \frac{1}{N} \sum_{i=1}^{N} X_{Mi} - \frac{1}{N} \sum_{i=1}^{N} X_{Oi}. \tag{2}$$

The *ubRMSD* is calculated as

$$ubRMSD = \left[\frac{1}{N} \sum_{i=1}^{N} \left[\left(X_{Mi} - \overline{X_M} \right) - \left(X_{Oi} - \overline{X_O} \right) \right]^2 \right]^{0.5}. \tag{3}$$

To indicate whether the modeled variability is greater than or less than the observed variability, the *ubRMSD* is multiplied by the sign of the difference in the modeled and observed standard deviations, such that

$$ubRMSD_2 = ubRMSD * \left(\sigma_M - \sigma_O \right) / \left| \sigma_M - \sigma_O \right|, \tag{4}$$

where σ_M and σ_O are the modeled and observed standard deviations, respectively. The bias and the $ubRMSD_2$ are normalized (denoted by subscript N) by the observed standard deviation to make their absolute values comparable among different variables:

$$bias_N = bias / \sigma_O, \tag{5}$$

$$ubRMSD_N = ubRMSD_2 / \sigma_O. \tag{6}$$

On each target diagram, the $bias_N$ between estimated and reference values is plotted on the Y-axis and the $ubRMSD_N$ is plotted on the X-axis. The radial distance from the origin to each data point is the normalized Root Mean Square Difference ($RMSD_N$).

To provide a more robust assessment than would be possible with a single metric, we used both the model skill from Willmott (1981) and the target diagram statistics from Jolliff et al. (2009) to assess model performance. To provide a succinct method to evaluate and report the accuracy of a large number of comparisons, we established a standardized set of cutoff values for both the skill scores and target statistics.

A range of reported model skill scores has been used to indicate model accuracy in predicting water level,

salinity and currents (e.g., Li et al. 2005; Warner et al. 2005a; Cheng et al. 2009; Liu et al. 2009; Ma et al. 2011; Zorndt et al. 2012; Bever and Harris 2014; Bever and MacWilliams 2013; Elias and Hansen 2013; MacWilliams and Gross 2013). Skill scores between 0.85 and 0.99 have been used to show that a model accurately predicts the water surface elevation. Skill scores of >0.80 have been used to show accurate modeling of salinity, with scores between 0.63 and 0.83 representing "reasonable" values. Skill scores >0.75 have been used to indicate accurate modeling of current speed or velocity, with scores between 0.68 and 0.90 indicating "acceptable" model predictions. Based on these values derived from the literature, we set thresholds for each variable. Above these thresholds, model predictions were considered accurate or acceptable, and below these thresholds, model predictions indicated poor agreement with observations (Table 1). Since the published literature did not indicate threshold skill scores for the accuracy of tidal flow predictions, the thresholds set for water level were also used for tidal flows.

The thresholds for classifying model performance using the target statistics were based on the distance from the origin of target diagram plots. The accuracy classification of the model was determined by circles of different radii from the origin of the target diagram (Table 1). The thresholds applied to classify the accuracy of the model predictions based on the target statistics were the same for each variable. Predictions falling outside a radius of 1 were classified as indicating poor agreement between the model predictions and the observed data. In this way, we classified as acceptable (or better) all predictions inside a radius of 1 that showed the model performed better than simply assuming the mean of the observations.

Observations Used in Model Evaluation

We compared model predictions to observations of water level, tidal flows, current speeds, and salinity at all stations throughout the estuary where data were available during the simulation period both graphically and by using the quantitative metrics described previously (Figure 1; see also Appendix A). Comparisons were made at 137 locations, using more than 5.6 million observations made during the 3-year analysis period. Continuous observations of water level were available from four stations in SF Bay and 30 stations in the Delta. Flow observations were available from six flow monitoring stations in the Delta.

Vertical profiles of horizontal velocity were collected using ADCPs at five locations in Suisun Bay during 1994, four locations during 1995, and three locations in 1996 (Burau et al. 1998; Kimmerer et al. 1998, 2002), with multiple deployments per year at some locations. We compared ADCP data with model predictions using two approaches. First, the predicted and observed depth-averaged current speeds were compared using the cross-correlation procedure described previously. Second, vertical profiles of horizontal velocity at the peak of each flood and each ebb tide were used to calculate the average peak

Table 1 Thresholds used for evaluating model accuracy based on model skill (Willmott 1981) and the radii of circles on the target diagram (Jolliff et al. 2009)

Model accuracy		Water level	Flow	Salinity	Current speed
Skill accuracy	Accurate	>0.975	>0.975	>0.85	>0.9
	Acceptable	0.95 – 0.975	0.95 – 0.975	0.7 – 0.85	0.8 – 0.9
	Poor agreement	<0.95	<0.95	<0.7	<0.8
Target accuracy	Very accurate	0.0 - 0.25			
	Accurate	0.25 - 0.5			
	Acceptable	0.5 - 1.0			
	Poor agreement	> 1.0			

flood and peak ebb tidal vertical profiles of horizontal velocity from observations and model predictions.

Salinity data were available from several sources. Continuous salinity observations were available from 38 monitoring stations, four of which had salinity measurements at two depths, and 26 of which were in the Delta. A series of vertical salinity profiles was taken at 11 stations through Suisun Bay (Entrapment Zone Study Salinity Station on Figure 1) to develop eight salinity transects during spring tides on April 27 and 28, 1994, and nine salinity transects during neap tides on May 17 and 18, 1994 (Burau et al. 1998; Kimmerer et al. 1998). The salinity transects were collected west to east over a 2-hour period at 3- to 5-hour intervals (Figure 1). The vertical salinity profile predicted by the model was saved at each sampling time and location to allow modeled and observed salinity to be compared directly.

Vertical salinity profiles at 1-m vertical resolution were collected as part of the USGS monitoring program at 38 sampling locations (Figure 1) along a 145-km transect from South San Francisco Bay to the western Delta (USGS 2013; Burau et al. 1998). We compared observed and predicted salinity at each sampling time and location for each of the 32 complete transects during the simulation period.

X2 and the Low-Salinity Zone

For each day, we calculated X2 as the distance from the Golden Gate to the location where the predicted daily-averaged near-bed salinity was 2 psu. We used the daily-averaged near-bed salinity along two transects following the axis of the estuary from the Golden Gate to the Sacramento–San Joaquin Delta (Figure 3) to calculate X2. For X2 > 75 km, we calculated X2 as the average of the distance along the Sacramento River and San Joaquin River transects (Figure 3).

X2 is often estimated using surface salinity from a small set of fixed observation stations that are typically near shore (e.g., Jassby et al. 1995; CDEC 2013), which introduces several sources of error into X2 estimation. First, the use of surface salinity requires an assumption about the amount of stratification.

Jassby et al. (1995) and Monismith et al. (2002) both assumed that the bed salinity is 2.0 psu where the surface salinity is 1.76 psu, whereas the current operational calculation of X2 (CX2) used to manage reservoir releases and Delta operations assumes the bed salinity is 2.0 psu when the surface salinity is 1.34 psu (CDEC 2013). Since stratification is not constant, both of these assumptions introduce error into X2 estimation. Second, the spatial interpolation between a small number of fixed stations requires an assumption about the salinity gradient between stations. Third, these approaches also assume a laterally-uniform salinity field, since near-shore stations are used to estimate salinity gradients in the channel.

The salinity profile data along the axis of the estuary collected by the USGS (Figure 1) provide one of the most direct ways to estimate X2 directly from observations, and can be used to evaluate the assumptions inherent in the X2 estimates that are based on the surface salinity observations discussed above. Though X2 is typically defined as a daily-averaged value, the USGS observations allow the instantaneous position of the near-bed 2.0-psu isohaline to be calculated based on direct observations of near-bed salinity along the axis of the estuary. The instantaneous position of the near-bed 2.0-psu isohaline was estimated from each of the 32 transects during the 3-year simulation period, and from a total of 239 transects between 1990 and 2013. The instantaneous position of the near-bed 2.0-psu isohaline was estimated from both the position of the observed and predicted 2.0-psu near-bed salinity interpolated between the stations, and from the position of the observed and predicted 1.76-psu surface salinity interpolated between the stations for each salinity transect. We compared the X2 estimates derived from observed surface and observed near-bed salinity to provide a measure of the error introduced into X2 estimates that are made from surface salinity observations based on an assumption of constant stratification. Additionally, we compared the position of the observed 2.0-psu near-bed salinity with the position of the 2.0-psu near-bed salinity predicted by UnTRIM at the same times that the transect data were collected to provide a measure of the error associated with the UnTRIM predictions of the instantaneous position

of the near-bed 2.0-psu isoahline. Since the daily-averaged position of the 2.0-psu near-bed salinity would be much more difficult to measure in the field, these comparisons of the instantaneous position of the near-bed 2.0-psu isohaline are one of the most direct ways to validate model predictions of X2 based on near-bed salinity. However, the instantaneous position of the near-bed 2.0-psu isohaline calculated using this approach cannot be compared directly to daily-averaged X2 calculations unless a correction is made to account for the tidal excursion between the instantaneous position of the near-bed 2.0-psu isohaline at the time of the observed salinity transect and the daily-averaged X2. This correction can be derived using the model predictions of daily-averaged X2 and the model prediction of the instantaneous position of the near-bed 2.0-psu isohaline at the USGS sampling locations and times.

For each day during the simulation period, the spatial extent of the low-salinity zone (LSZ) was calculated and related to the predicted X2. For this analysis, we defined the LSZ as the region where the depth-averaged salinity was between 0.5 and 6. The percentage of each day during which the modeled depth-averaged salinity within each cell was within this range indicated the daily tidal excursion of the LSZ. The area and volume of all grid cells from San Pablo Bay through the western and central Delta (Figure 1) with daily-averaged depth-averaged salinity between 0.5 and 6 were then used to represent the area and volume of the LSZ.

X2 Autoregressive Equation

Several relationships have been proposed to estimate X2 as an autoregressive function of Delta outflow (Jassby et al. 1995; Monismith et al. 2002; Gross et al. 2009). The form of the equation proposed by Monismith et al. (2002) is consistent with theoretical predictions of salinity intrusion and is therefore used here:

$$X2(t) = \alpha(t)X2(t-1) + (1-\alpha(t))\beta Q(t)^{\gamma} \quad (7)$$

Here, t is time in days, $Q(t)$ is the current day's net Delta outflow in $m^3\ s^{-1}$, $\alpha(t)$ is the weight between the autoregressive term and the flow term, β is a scaling coefficient, and γ is the power law exponent that indicates the sensitivity of the salinity field to flow variability. The steady form of Equation 7 is

$$X2_{steady} = \beta Q_{steady}^{\gamma} \quad (8)$$

We take α to vary linearly with Delta outflow as suggested by Monismith et al. (2002), being bounded in the range [0,1], i.e.,

$$\alpha(t) = min\left[1, max\left[0, \alpha_0\left(mQ(t)+b\right)\right]\right] \quad (9)$$

where m is the slope and b the intercept of the linear relationship determined in the first step of fitting, and α_0 is a parameter that will be determined in the second step.

The data used to fit the parameters m and b were the rates of convergence of X2 to steady-state X2 for each of nine steady flow scenarios simulated using the UnTRIM Bay–Delta model (Kimmerer at al. 2013). For these scenarios, we assume that X2(t) is given by

$$X2(t) = X2_{steady} + Ae^{-\frac{t}{T}} \quad (10)$$

where X2$_{steady}$ is the steady-state X2 value reached at the end of a steady flow simulation, T is the time scale for the approach to steady state, and A is a scaling parameter. Following Monismith et al. (2002), m and b in Equation 9 are related to T by

$$mQ_{steady} + b = \frac{T-1}{T} \quad (11)$$

The data set used to determine the remaining parameters (α_0, β, and γ) consists of daily Delta outflow and corresponding X2 values calculated from UnTRIM Bay–Delta model results ("UnTRIM X2") for each day during the 3-year simulation period.

We constructed probabilistic model of Equation 7 in WinBUGS version 1.4 (Spiegelhalter et al. 2003), and performed Bayesian inference in two steps. First, we determined m and b using X2 estimates from the nine steady flow scenarios (Kimmerer et al. 2013) and uninformative prior probability distributions:

$$m \sim \min\left[0, \mathcal{N}\left(-10^{-4}\,\text{s m}^{-3}, 10\,\text{s}^2\text{m}^{-6}\right)\right];\ b \sim \min\left[1, \mathcal{N}\left(1, 10\right)\right] \tag{12}$$

where $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2. Then we determined the remaining parameters (α_0, β, and γ) using the X2 predictions from UnTRIM for the 3-year simulation period and the following uninformative priors for the parameters:

$$\alpha_0 \sim \max\left[0, \mathcal{N}\left(1,1\right)\right];\ \beta \sim \max\ 0, \mathcal{N}\left(300, 10^6\right);\ \gamma \sim \mathcal{U}\left(-0.4, -0.1\right) \tag{13}$$

where $\mathcal{U}(c, d)$ is a uniform distribution with a minimum value of c and maximum value of d. Each step of the fitting was run with triplicate Markov chains of 2,000 samples per chain after 10-fold thinning (i.e., retaining only every tenth sample in each chain), including a burn-in period of 50 (post-thinning) samples. This resulted in 1,950 retained samples per chain that were used in forming posteriors. Statistics of the first and second halves of the retained samples were compared to check that both the burn-in period and the number of samples were adequate for convergence. We used Gelman–Rubin statistics (Gelman et al. 2004), plots of autocorrelation, and time series of samples to check model convergence. The substantial computation time required by each forward model run in WinBUGS limited the analysis to a relatively small number of samples. As an additional check on the WinBUGS results, we also fit the parameters in Matlab using the 'nlinfit' function used by Gross et al. (2009) in an earlier X2 analysis. This was done both for a constant-α equation of the form used by Monismith et al. (2002) and for the new variable-α equation (Equation 7) developed in this study.

The new form of the equation (Equation 7) suggests a specific concept of antecedent flow directly applicable to X2. The antecedent flow, $Q_{ant}(t)$, for a given day is the flow that, if used in the steady form of the X2 equation, would result in the X2 estimated for that day using Equation 7. Therefore, antecedent flow can be calculated by simply equating the steady form of the X2 equation with the unsteady form with Q_{steady} replaced with $Q_{ant}(t)$. This gives

$$Q_{ant}(t) = \left[\alpha(t)\beta^{-1}X2(t-1) + \left(1-\alpha(t)\right)Q(t)^{\gamma}\right]^{\frac{1}{\gamma}} \tag{14}$$

We can express $Q_{ant}(t)$ as a weighted time average of preceding flows by recursively substituting the equations for X2 on past days into the above equation. Performing this substitution for the previous M days and discarding any remaining terms that depend on flows before day t-M gives the following expression for $Q_{ant}(t)$:

$$Q_{ant}(t) = \left[\left(1-\alpha(t)\right)Q(t)^{\gamma} + \sum_{t'=t-M}^{t-1}\left(\prod_{t''=t'+1}^{t}\alpha(t'')\right)\times\left(1-\alpha(t')\right)Q(t')\gamma\right]^{\frac{1}{\gamma}} \tag{15}$$

14

RESULTS

Evaluation of Model Accuracy

Appendix A presents a comprehensive set of comparisons between observations and model predictions. For each comparison, the coefficient of determination, model skill, and target statistics are tabulated to allow for an objective evaluation of model accuracy based on the accuracy thresholds shown in Table 1.

An example comparison for water level in at the Sacramento River North of the Delta Cross Channel station (WGA on Figure 1) shows that the model accurately predicts water levels on a tidal time scale over 15 days (Figure 5A). The comparison of tidally averaged water level for the full analysis period shows that the model accurately predicts spring–neap effects as well as non-tidal forcing such as freshwater flow and storm surge (Figure 5B). The effect of periods of high flow on water levels during winters is clearly evident in the tidally averaged

comparison (Figure 5B). Cross-correlation with lags removed (MacWilliams and Gross 2013) shows a largely tight, nearly 1:1 relationship between observations and model output (Figure 5C). The model accurately predicted water levels on tidal time scales during all freshwater flow conditions ($r^2 = 0.994$ and skill $= 0.997$). Figures with a similar format are used for all time–series comparisons shown below and summarized in tabular form in Appendix A. The accurate prediction of water levels achieved at most stations (Appendix A) indicates that tides were accurately propagated from the model boundary at the Pacific Ocean through the entire estuary.

The eight salinity transects collected on April 27 and April 28, 1994 show the evolution of the salinity field in Suisun Bay during two tidal cycles of a spring tide (Figure 6). Salinity along the transects varied from more than 20 psu to less than 0.5 psu and the model accurately predicted both the salinity gradients along the transects and the degree of strati-
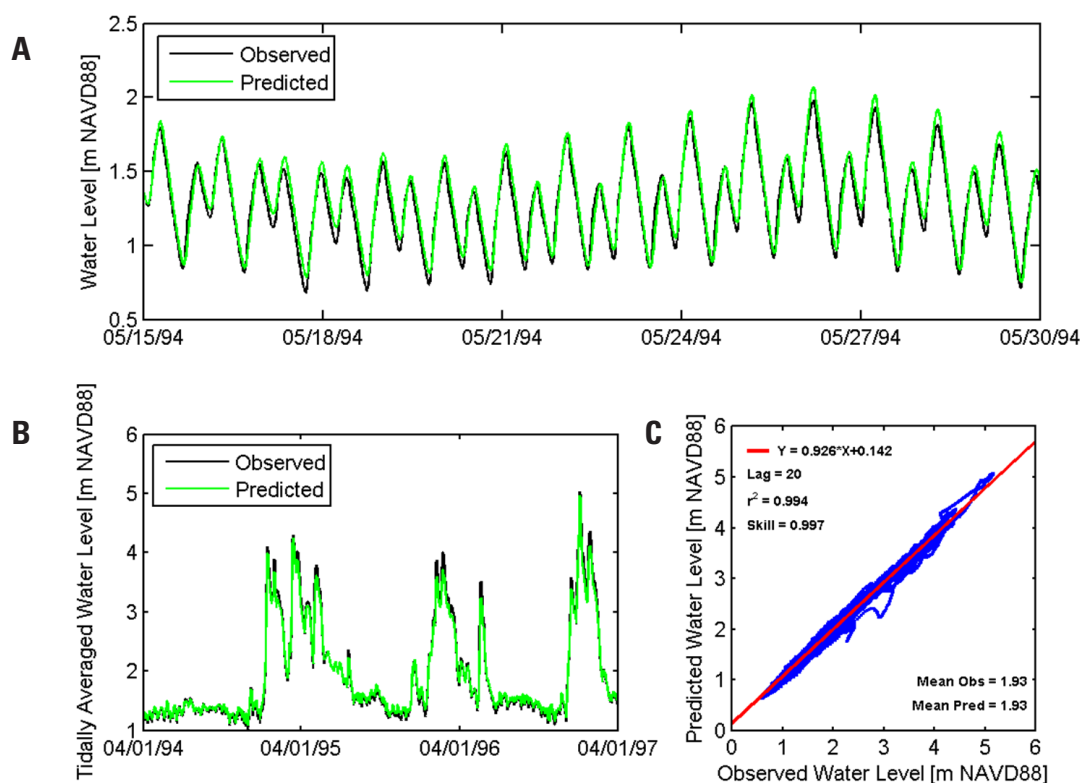

Figure 5 Observed and predicted water level at the Sacramento River North of Delta Cross Channel USGS station (WGA on Figure 1) during the 1994–1997 simulation period
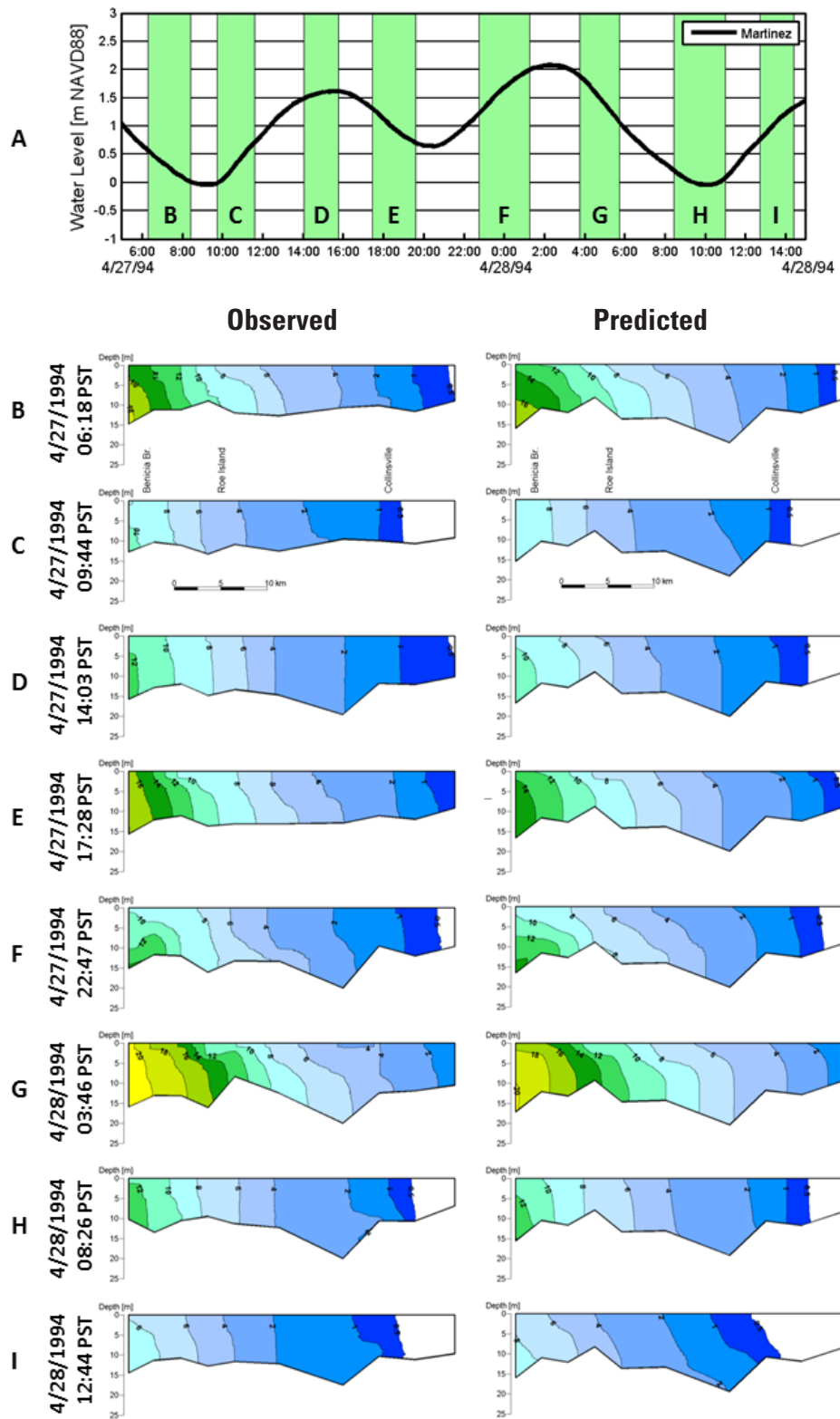
15

Figure 6 Observed and predicted salinity profiles during spring tide transects on April 27 and April 28, 1994 during eight periods indicated on top panel (**A**) by green bars (**B** through **I**).

fication (Figure 6). Both the observed and predicted salinity show that the isohalines were advected over 5 to 10 km during a flood or ebb tide and salinity remained relatively well-mixed during most of this period. The strongest stratification was evident near higher high water (Figure 6G), and appears to be the result of the relatively strong longitudinal salinity gradients that developed during flood tide and reduced vertical mixing near slack water. The model accurately predicted the timing, location, and magnitude of stratification throughout the spring tide transect period. The average error—measured as the difference between the observed and predicted salinity at each vertical and horizontal sampling point along each of the eight spring tide transects—was -0.12 psu. The r^2 values computed for each transect ranged from 0.986 to 0.994, and the model skill ranged from 0.980 to 0.997 (Table 2).

The nine salinity transects collected on May 17 and May 18, 1994 show the evolution of the salinity field in Suisun Bay during two tidal cycles of a neap tide (Figure 7). Salinity transects collected during neap tides indicated somewhat greater stratification in both observed and predicted profiles than during the spring tide period. Both observed and predicted stratification persisted through ebb tides and through the weaker flood tide on May 18 (Figure 7G). The timing, location, and magnitude of stratification was also predicted accurately throughout the neap tide transect period. The average error in salinity for the nine neap tide transects was 0.03 psu, the r^2 values ranged from 0.965 to 0.987, and the model skill ranged from 0.987 to 0.996 (Table 3).

The comparison of predicted salinity profiles to observed salinity data collected at stations along the axis of the estuary (USGS Water Quality Sampling Station on Figure 1) demonstrate that the model accurately represented the position of isohalines during periods of low and high Delta outflow (Figure 8). On October 26, 1994, (Dayflow outflow = 125 m^3s^{-1}) the observed and predicted salinities ranged from more than 32 psu near the bed in Central Bay to less than 1 psu near Rio Vista at the upstream end of the transect. On January 18, 1995 (Dayflow outflow = 3,714 m^3s^{-1}), observed and predicted salinities ranged from more than 28 psu near the bed in Central Bay to less than 0.5 psu east of Carquinez Bridge. On January 18, 1995, both the observed and

Table 2 Comparisons between model output and salinity data from longitudinal salinity transects during the Entrapment Zone study during spring tides (Figure 6) for UnTRIM (this study) and TRIM3D (Gross et al. 2009)

|	UnTRIM Bay–Delta Model				TRIM3D San Francisco Bay Model		
Date and time	Average error (psu)	Standard error (psu)	r^2	Skill	Average error (psu)	Standard error (psu)	r^2
4/27/1994 06:18	0.07	0.79	0.988	0.995	−1.63	0.82	0.980
4/27/1994 09:44	−0.02	0.37	0.992	0.997	−1.48	0.98	0.961
4/27/1994 14:03	−0.72	0.57	0.994	0.987	−1.58	0.92	0.973
4/27/1994 17:28	−0.49	1.05	0.986	0.988	−0.64	0.63	0.987
4/27/1994 22:47	0.44	0.60	0.980	0.992	−1.55	1.07	0.938
4/28/1994 03:46	−0.09	1.02	0.987	0.994	−1.54	0.84	0.985
4/28/1994 08:26	0.03	0.48	0.992	0.997	−1.78	0.97	0.957
4/28/1994 12:44	−0.20	0.38	0.988	0.995	−1.48	1.09	0.916
Average spring tide	−0.12	0.66	0.988	0.993	−1.46	0.92	0.962
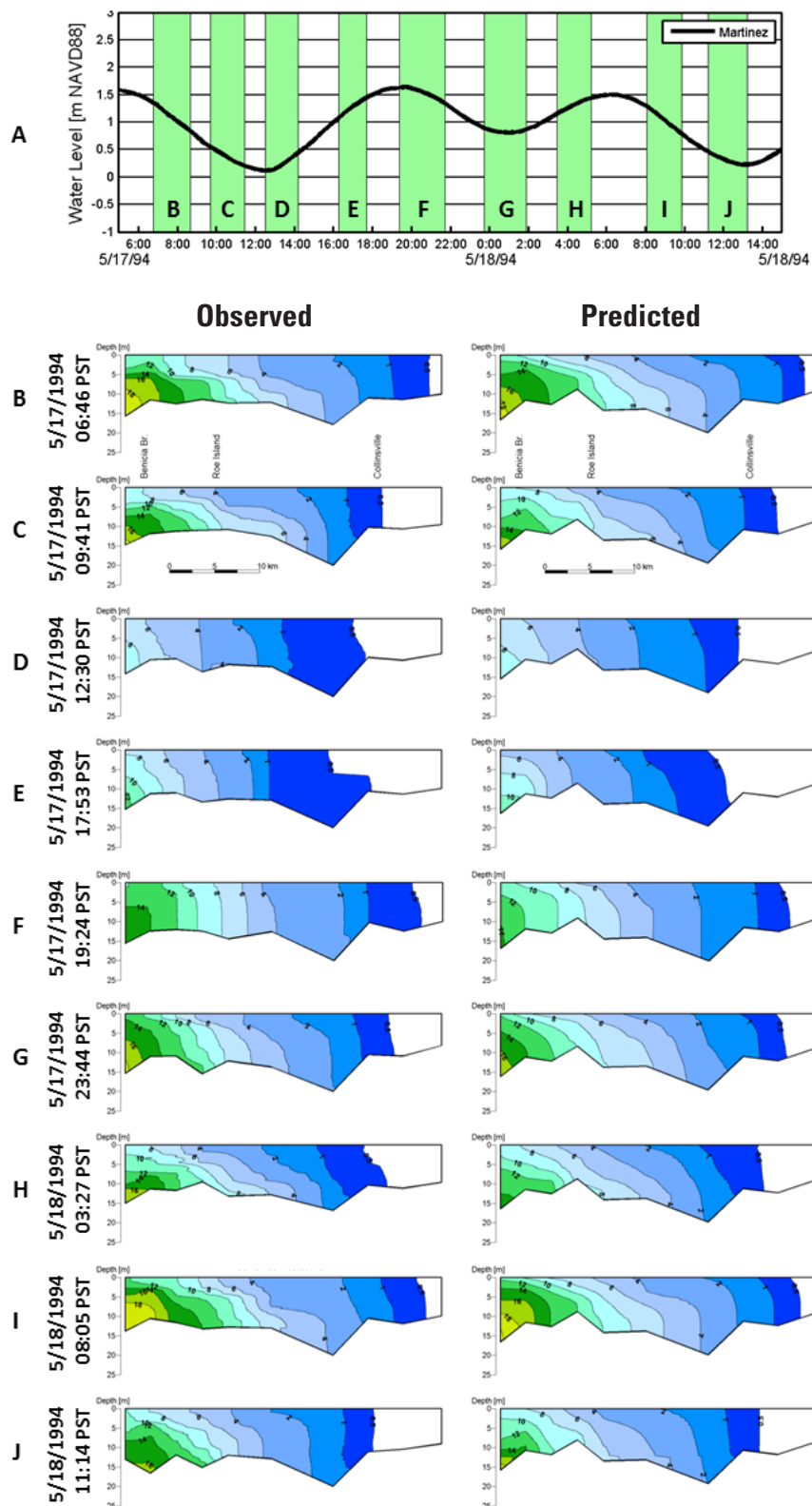
Figure 7 Observed and predicted salinity profiles during neap tide transects on May 17 and May 18, 1994, during nine periods indicated on top panel (**A**) by green bars (**B** through **J**)

Table 3 Comparisons between model output and salinity data from longitudinal salinity transects during the Entrapment Zone study during neap tides (Figure 7) for UnTRIM (this study) and TRIM3D (Gross et al. 2009).

Date and time	UnTRIM Bay–Delta Model				TRIM3D San Francisco Bay Model		
	Average error (psu)	Standard error (psu)	r^2	Skill	Average error (psu)	Standard error (psu)	r^2
5/17/1994 06:46	0.57	1.10	0.965	0.989	–0.70	1.33	0.948
5/17/1994 09:41	0.12	0.90	0.975	0.992	–1.25	1.52	0.921
5/17/1994 12:30	0.13	0.36	0.986	0.996	–0.76	0.58	0.974
5/17/1994 17:53	–0.37	0.69	0.982	0.987	–0.97	1.02	0.981
5/17/1994 19:24	–0.58	0.72	0.987	0.991	–0.50	0.71	0.981
5/17/1994 23:44	0.19	0.89	0.983	0.993	–0.73	0.68	0.988
5/18/1994 03:27	0.15	0.77	0.979	0.993	–1.59	1.66	0.925
5/18/1994 08:05	0.13	0.91	0.981	0.994	–1.15	1.44	0.949
5/18/1994 11:14	–0.07	0.79	0.979	0.993	–1.26	1.33	0.939
Average neap tide	0.03	0.79	0.980	0.992	–0.99	1.14	0.956

predicted salinity stratification were more than 20 in Central Bay. Based on the accuracy thresholds shown in Table 1, the model accurately predicted all 32 of the salinity transects (Table 8 in Appendix A), with a model skill between 0.981 and 0.999.

Salinity 0.9 m above the bed at Point San Pablo (PSP on Figure 1) was accurately predicted on tidal time scales (Figure 9A). Both observed and predicted tid-ally-averaged salinity (Figure 9B) increased through 1994, falling rapidly because of high outflows in January and March 1995, and increasing from May 1995 through October 1, 1995. A similar seasonal pattern was evident in 1996 and 1997. Further land-ward at Mallard Island (MAL on Figure 1) the pre-dicted surface salinity was also similar to observed surface salinity over the 3-year comparison period (Figure 10). The accurate prediction of salinity at most stations under a very wide range of Delta out-flows (see "Discussion" and Appendix A) indicates that the model is accurately predicting both the intrusion of oceanic salt into the estuary, and the transport of salts derived from agricultural runoff through the Delta into the estuary.

X2 and the Low-Salinity Zone

The predicted daily X2 value and area of the LSZ varied over a wide range of values during the 3-year simulation period (Figure 11). The relationships between X2 and the area and volume of the LSZ are not monotonic (Figure 12A, 12B). Local minima of LSZ area occurred at constrictions at Carquinez Strait (X2 ∼ 55 km) and in the western Delta (X2 ∼ 86 km). Variability generally increased with decreasing X2, and was greatest west of Carquinez Strait. The aver-age depth of the LSZ was roughly inversely related to area (Figure 12C), and was highest when X2 was in Carquinez Strait or the western Delta.

The instantaneous position of the near-bed 2.0-psu isohaline was estimated from both the observed salin-ity profile data collected by the USGS (e.g., Figure 8) and from the corresponding model predictions at the same locations and times that the USGS mea-sured salinity. The observed and predicted location of the near-bed 2.0-psu ishohaline were estimated from the near-bed salinity and from the position of the 1.76-psu surface salinity. For the 32 USGS salin-
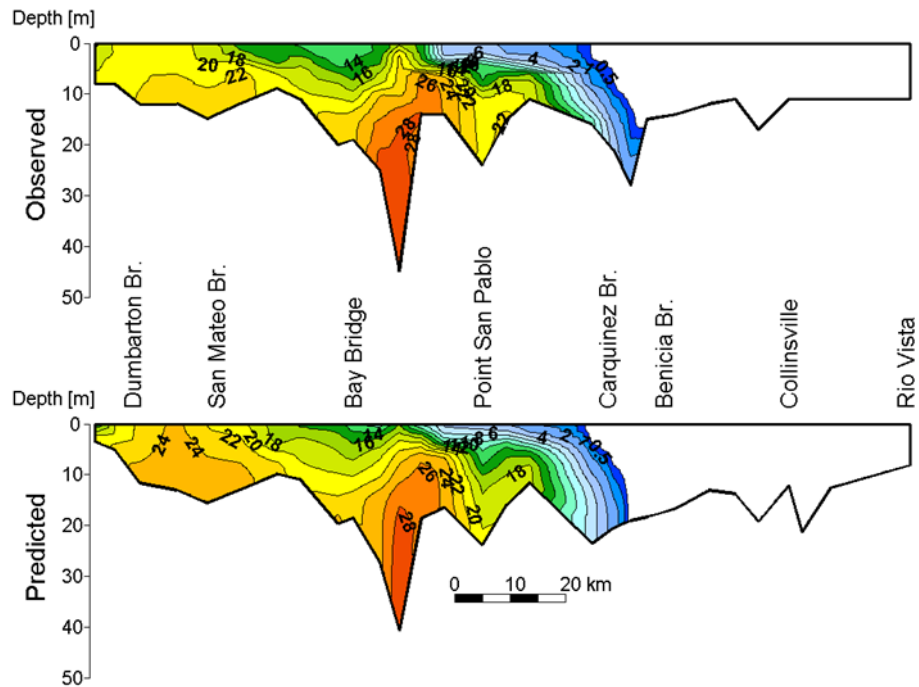
A) October 26, 1994
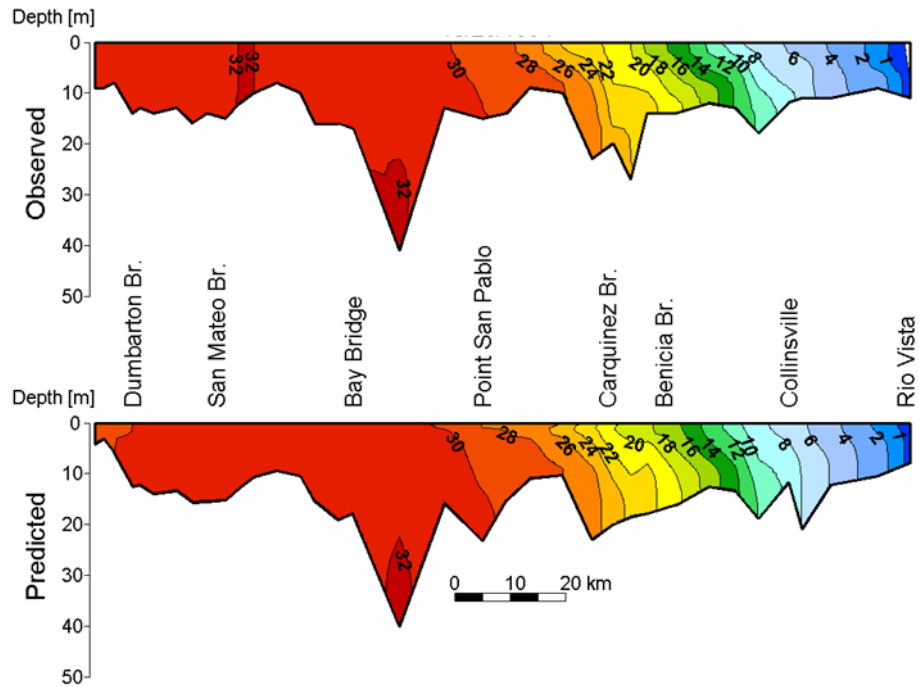
B) January 18, 1995

Figure 8 Observed and predicted salinity profiles from transects interpolated along the axis of the San Francisco Estuary on (**A**) October 26, 1994 and (**B**) January 18, 1995
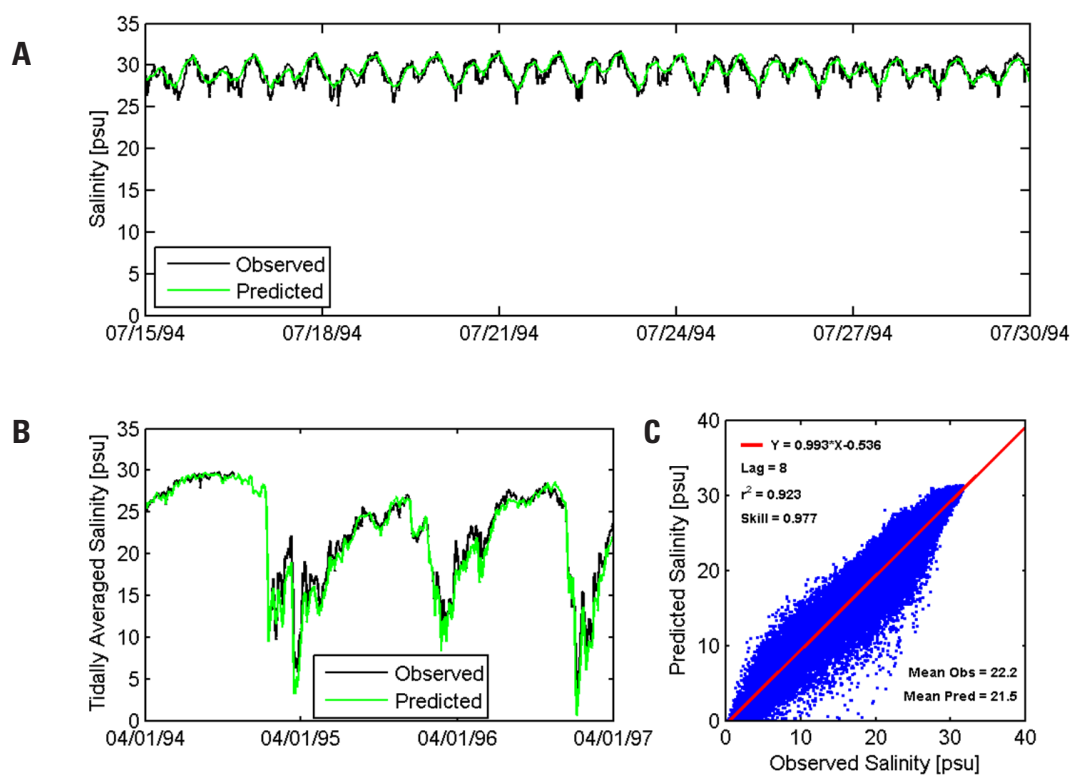
Figure 9 Observed and predicted salinity from the lower sensor at Point San Pablo (PSP on Figure 1) during the 1994–1997 simulation period
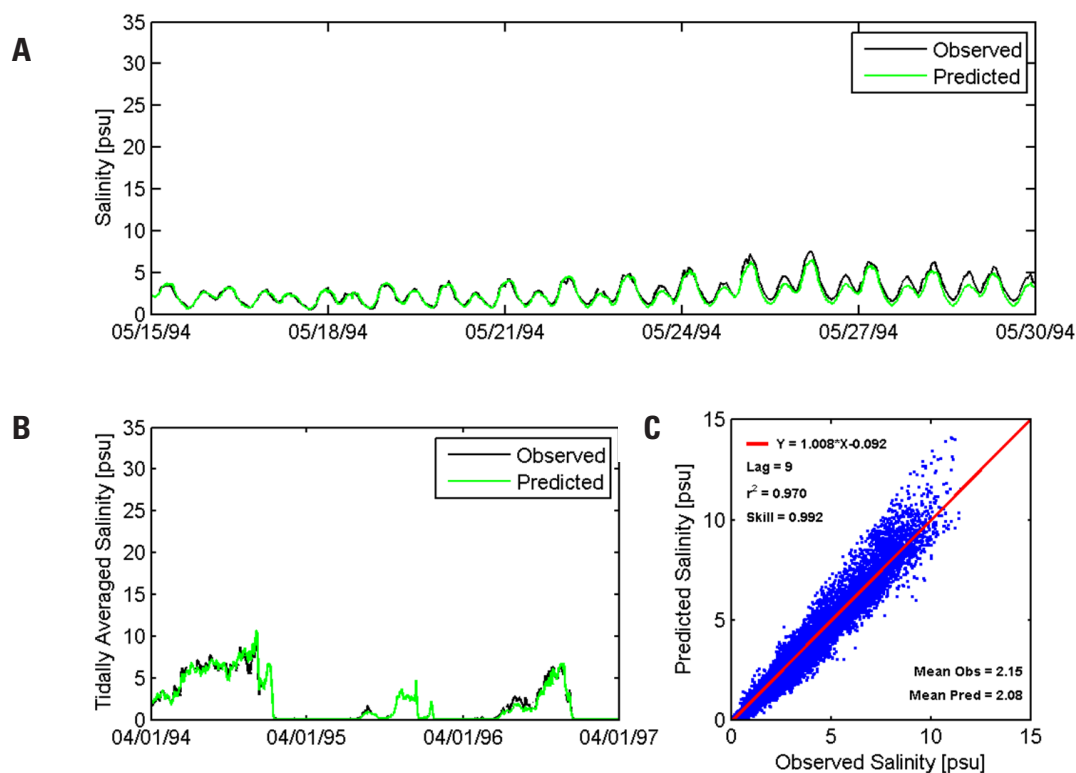

Figure 10 Observed and predicted salinity from the surface sensor at the Sacramento River near Mallard Island station (MAL on Figure 1) during the 1994–1997 simulation period
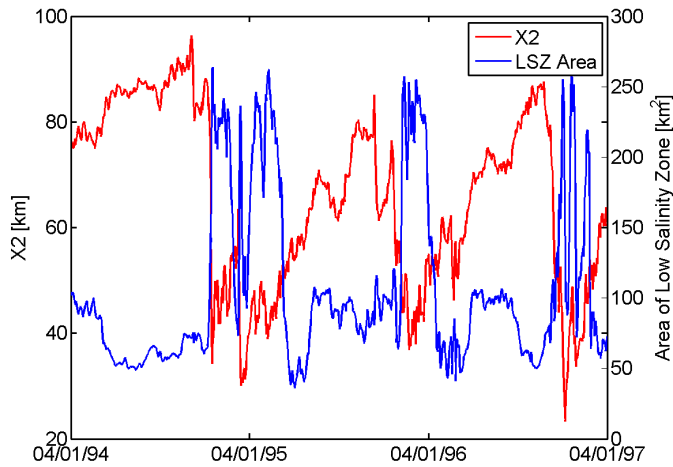
21

Figure 11 Predicted X2 and LSZ area for each day during the 1994–1997 model simulation
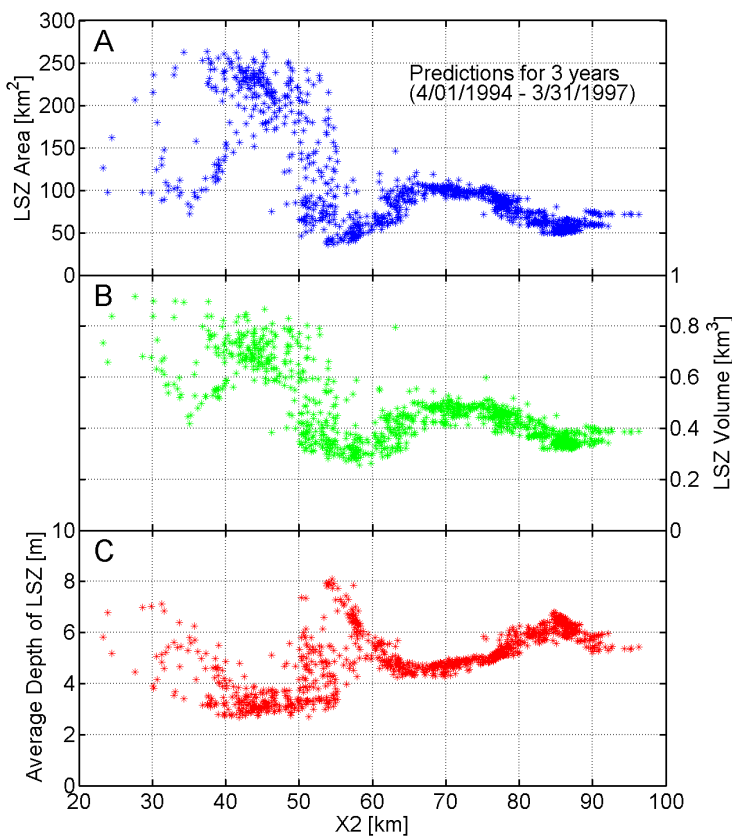

Figure 12 Relationship between daily X2 and the daily-averaged (**A**) area, (**B**) volume, and (**C**) depth of the low-salinity zone (LSZ)

ity profiles during the analysis period, the UnTRIM predictions of the instantaneous position of the near-bed 2.0-psu isohaline more accurately matched the observed position of the near-bed 2.0-psu isohaline than the corresponding estimates derived from the observed 1.76-psu surface salinity (Figure 13 and Table 4). Both the model predictions and estimates based on observed surface salinity showed more scatter when X2 was less than 70 than when X2 was greater than 70. For X2 > 70, the mean absolute value of the error in the UnTRIM X2 predictions was calculated to be 1.22 km, while the mean absolute value of the error for the X2 estimates derived from observed surface salinity was 1.64 km (Table 4). For X2 < 70, the mean absolute value of the error in UnTRIM X2 predictions was estimated to be 4.17 km, while the mean absolute value of the error for the X2 estimates derived from observed surface salinity was 6.82 km. The larger errors associated with lower X2 may be associated either with uncertainty in the flow estimates during high flows or the transient nature of the salinity field during high Delta outflows. The error estimates associated with using the 1.76-psu surface salinity to estimate X2 calculated for the 1990–2013 data collection period (Table 4) are similar to those for the 3-year simulation period. These comparisons indicate that over the full range of X2, the instantaneous position of the 2.0-psu isohaline calculated from the predicted near-bed salinity in UnTRIM consistently has less error than is introduced by the common assumptions used to calculate daily-averaged X2 from surface salinity.

X2 Autoregressive Equation

A time scale for the approach to a steady state X2 value was estimated using Equation 10 for each of the nine steady flow scenarios described by Kimmerer at al. (2013). The WinBUGS analysis gave mean values for m and b (Equation 9) of $-1.12 \times 10^{-4}\,\mathrm{s\,m^{-3}}$ and 0.977, respectively, with 95% credible intervals reported in Table 5. After the posteriors of m and b were determined, we estimated the remaining parameters, which had small credible intervals (Table 5). The time series of predictions from this model had a 95% credible interval that included 95.6% of the UnTRIM predictions (Figure 14) and typically spanned 12 km.
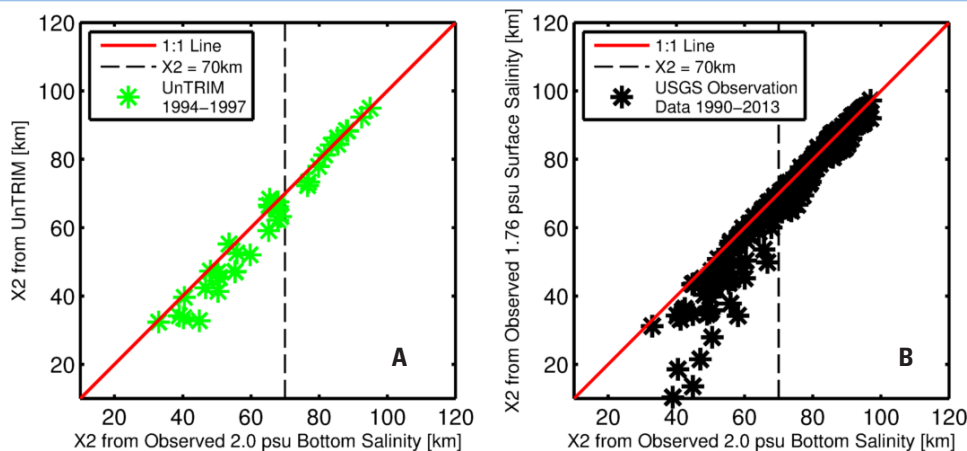
Figure 13 Comparison of X2 estimated from the position of observed 2.0 psu bottom salinity from USGS salinity transects with X2 predicted by the UnTRIM Bay-Delta model (**A**) and the X2 estimated from the position of observed 1.76 psu surface salinity from the USGS salinity transects (**B**)

The parameters α, β, and γ, and error metrics of X2 predicted by various equations compared with UnTRIM X2 are substantially different among the constant-α and variable-α equations and published equations (Table 6, Figure 15). The new variable-α equation is the most accurate, and is a significant improvement over the constant-α equation fit to the same UnTRIM X2 values. The X2 values for various steady flows are similar among the models for moderate Delta outflows but diverge widely at high outflow (Figure 16).

We evaluated Equation 15 for antecedent flow for M = 365 days so that a full year of daily flows was used to estimate antecedent flow for the simulation period (Figure 17). Because α is proportional to Delta outflow, the antecedent flow is similar to the daily outflow for high Delta outflow and very different from daily outflow for low Delta outflow. Using this form of antecedent flow in the steady X2 equation (Equation 8) is equivalent to using the unsteady equation (Equation 7). The UnTRIM X2 versus antecedent flow for each day of the simulation period is compared with the X2 values predicted by the steady forms of the regression equations in Figure 16.

Table 4 Estimates of mean absolute error, RMS error, and model skill calculated from the difference between the X2 estimated from the position of observed 2.0 psu bottom salinity from USGS salinity transects with X2 calculated from the UnTRIM Bay-Delta model and the X2 estimated from the position of observed 1.76 psu surface salinity from the USGS salinity transects. Error and skill estimates are shown for X2 > 70 and X2 < 70 based on the division shown on Figure 13.

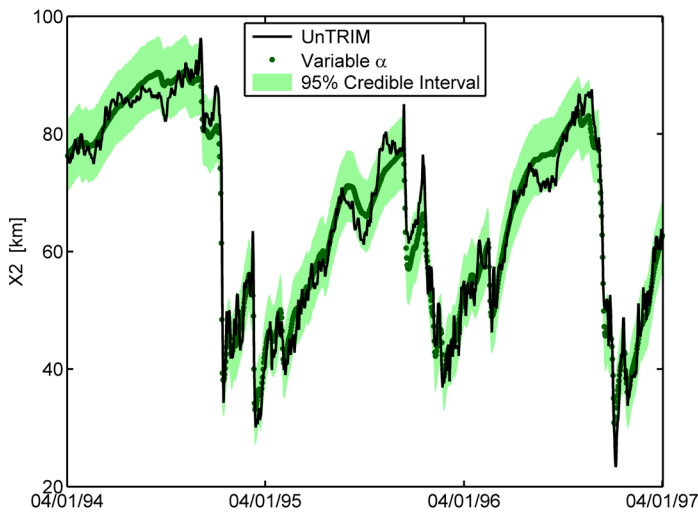| X2 range	X2 method	Evaluation period	Mean absolute error (km)	RMS Error (km)	Skill
X2 > 70	UnTRIM	1994–1997	1.22	1.85	0.979
	Surface salinity	1994–1997	1.64	2.14	0.958
		1990–2013	2.23	3.06	0.963
X2 < 70	UnTRIM	1994–1997	4.17	5.24	0.946
	Surface salinity	1994–1997	6.82	9.94	0.863
		1990–2013	5.96	9.02	0.839

Figure 14 X2 calculated from UnTRIM Bay–Delta model results and mean and 95% credible interval for Bayesian estimation of the parameters in Equation 7
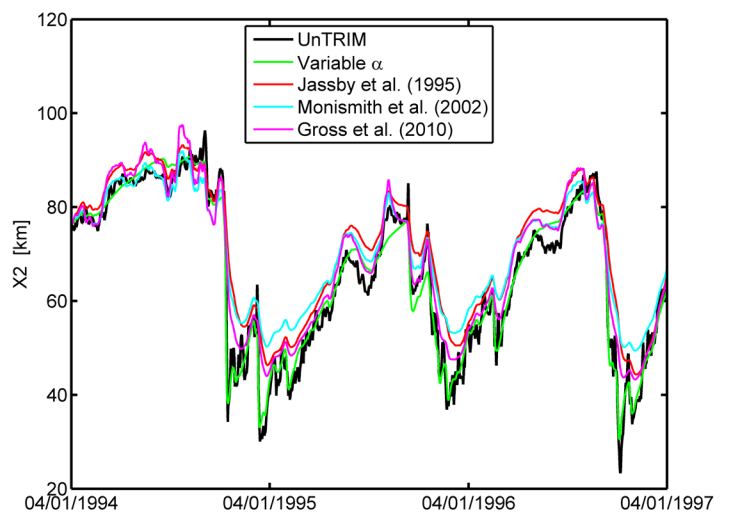

Figure 15 X2 calculated from UnTRIM Bay–Delta model results, the variable-α regression, and the regressions developed by Jassby et. al (1995), Monismith et al. (2002), and Gross et al. (2009)

Table 5 Statistical model of X2 vs. outflow with variable-α

Parameter	Mean	Median	Std. Dev.	2.5%	97.5%
m	-0.0001116	-0.0001116	8.221×10^{-7}	-0.0001133	-0.0001100
b	0.9773	0.9773	0.0002615	0.9768	0.9778
α_0	1.025	1.026	0.00115	1.023	1.027
β	284.1	284.4	5.867	271.6	294.9
γ	-0.2299	-0.2301	0.002887	-0.2352	-0.2236

Table 6 Parameter values and error metrics in predicting UnTRIM X2

Autoregressive equation	α	β	γ	RMS error (km)	r^2
Jassby et. al (1995)	NA[a]	NA[a]	NA[a]	7.33	0.806
Monismith et al. (2002)	0.919	167.5	-0.141	7.47	0.798
Gross et al. (2009)	0.910	210.0	-0.182	5.31	0.898
Constant-α (this study)	0.895	226.6	-0.202	4.17	0.937
Variable-α (this study)	Variable	284.1	-0.230	3.10	0.965

a. NA indicates "not applicable." The parameters are not relevant for the Jassby et al. (1995) autoregressive equation because it has a different form.
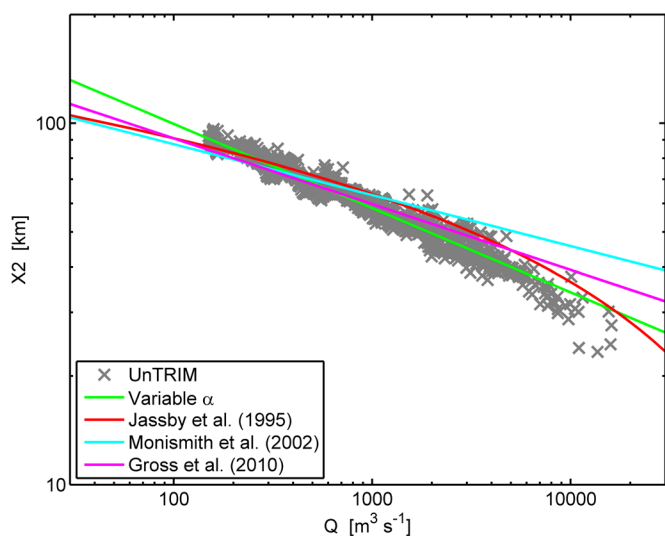
Figure 16 X2 predicted by UnTRIM and the steady forms of the variable-α regression, the regressions developed by Jassby et al. (1995), Monismith et al. (2002), and Gross et al. (2009). The relevant flow value for each daily X2 value in the UnTRIM data set is the antecedent flow on that day.

Figure 17 Daily Delta outflow and antecedent flow during the three-year UnTRIM simulation period

DISCUSSION

Evaluation of Model Accuracy

The UnTRIM San Francisco Bay–Delta model was the first 3-D hydrodynamic model to include the entire region from the Pacific Ocean through SF Bay and the entire Sacramento–San Joaquin Delta. The model calibration has been documented in numerous technical reports (e.g., MacWilliams et al. 2008, 2009; MacWilliams and Gross 2010; MacWilliams and Bever 2013). This paper is the first peer-reviewed publication that fully documents the model calibration and the improvements over the previously published TRIM3D model results (Gross et al. 2009).

Previous modeling studies in the San Francisco Estuary using the TRIM3D model (e.g., Gross et al. 2009) found that the geometric and resolution limitations of applying a structured Cartesian grid substantially affected hydrodynamic and salinity predictions in some regions of the estuary. In particular, most of the Delta in the TRIM3D application was represented as a pair of tidal lakes tuned to make tidal flows match observations. This likely influenced tid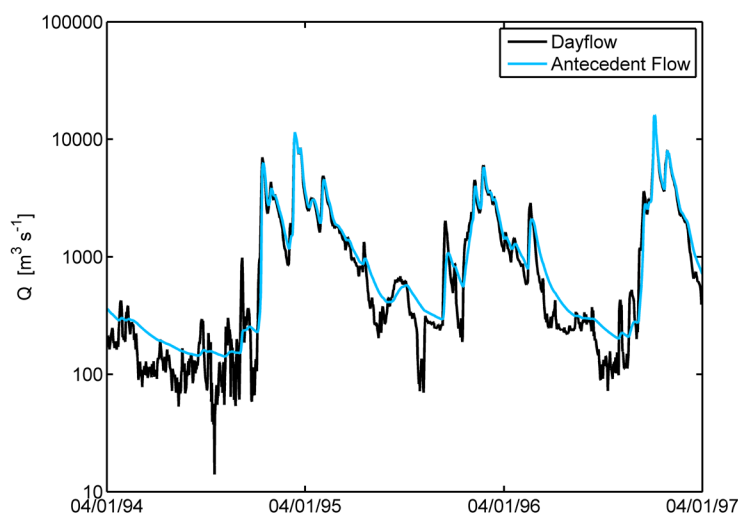al flows throughout the upper estuary, and, consequently, salinity patterns. Analyses of habitat for fish using TRIM3D (Kimmerer et al. 2009) and UnTRIM (Kimmerer et al. 2013) gave similar results for species in brackish to saline regions of the estuary, indicating agreement in the transport of salt at a daily- and depth-averaged scale. However, as will be described here, the current UnTRIM Bay–Delta model implementation more accurately predicts stratification and tidal variability of salinity than the TRIM3D San Francisco Estuary implementation presented by Gross et al. (2009).

Comparisons of the model calibrations between UnTRIM in this study and previous applications with TRIM3D (e.g., Gross et al. 2006, 2009) show a substantial reduction in average error and standard error and an increase in the coefficient of determination in the prediction of salinity on transects through Suisun Bay (Burau et al. 1998) using UnTRIM (Tables 2 and 3). In 14 of 17 comparisons with observations, the standard error was lower for UnTRIM than for TRIM3D, and the average error was lower for UnTRIM in 16 of 17 comparisons. The average error for the spring tide transects was

reduced from −1.46 to −0.12 psu and the average error for the neap tide transects was reduced from −0.99 to 0.03 psu using UnTRIM. The median ratios of standard and average error from UnTRIM to that from TRIM3D were 0.62 and 0.14 respectively. Thus, UnTRIM resulted in an increase in both the precision and the accuracy of the salinity predictions. Since many aspects of the numerical methods used in the TRIM3D model and the UnTRIM model are similar, much of the improvement in the salinity predictions from the UnTRIM implementation can be attributed to better grid resolution through Carquinez Strait and Suisun Bay than in the TRIM3D implementation. In addition, the TRIM3D model implementation used a "false delta" consisting of two rectangles to represent the majority of the Delta, whereas the unstructured and channel-aligned grid used in UnTRIM (Figure 4) allows for detailed representation of the bathymetry in narrow and sinuous channels and junctions in the Delta which could not be adequately resolved using a structured grid in TRIM3D. However, some improvements in accuracy may also result from more recent improvements to numerical methods used in UnTRIM, such as improved numerical methods for scalar transport (e.g., Casulli and Zanolli 2005).

Stow et al. (2009) highlighted a lack of rigorous assessment of physical models in the published literature and reported that the most widely used model skill assessment metric was simply a visual plot of the data and the predictions. The lack of rigorous assessment arises because the assessment of model accuracy depends on the observed variable, the uncertainty in the observed data, and the questions being investigated with the model (Stow et al. 2009). Using multiple quantitative metrics, as we did here, allows for both a more rigorous approach to quantify how well predicted values match the observations and a more objective method to characterize model accuracy.

We applied discrete thresholds (Table 1) for the model skill (Willmott 1981) and the target diagram statistics (Jolliff et al. 2009) to categorize the stations based on the accuracy to which the model predicted the observations. These evaluations yielded somewhat similar assessments of model accuracy (Table 7). Both metrics show that water level was more accurately pre-

dicted at a larger percentage of stations than salinity, but the exact number of stations within each accuracy classification was different for the two metrics. This highlights that both the model skill and the target statistics should be included in any rigorous and robust assessment of model performance.

These two assessment metrics show that predictions of water level, tidal flow, and current speed were almost all classified as accurate or very accurate, and comparisons of these variables at all but one station were at least acceptable during the 1994–1997 simulation period (Table 7). The model predictions of salinity were acceptable or accurate at 87% of the stations based on skill, and at 85% of the stations based on the target metric (Table 7). About 15% of the salinity stations fell within the lowest classification based on the target accuracy criteria. These poorly performing stations were generally at locations where the average salinity during the simulation period was less than 0.5 (Figure 18). These stations tend to have only small seasonal signals and can be significantly influenced by salinity contributions from agricultural drainage, which is not accurately known. Additionally, some observation data appear to be erroneous, which can result in error metrics that indicate poor agreement between model predictions and observations. Thus, in some cases, an indication of poor agreement can result from poor data quality rather than from poor model accuracy. The tables in Appendix A provide the accuracy metrics for all stations plotted on Figure 18.

The model assessment metrics applied here do not depend on the simulation year, so metrics can be compared between different time-periods, provided both periods are sufficiently long and span a wide enough range of conditions to thoroughly test model performance. To demonstrate the comparison of different years we compared the 1994–1997 simulation to a simulation spanning January 1 to December 31, 2012 (Figure 18, Table 7) during which observation data were available at more than twice as many continuous monitoring locations (MacWilliams and Bever 2013) than during the 1994–1997 simulation period (186 in 2012 compared to 78 in 1994–1997). The model assessment generally shows a higher percentage of stations classified as either "very accurate"
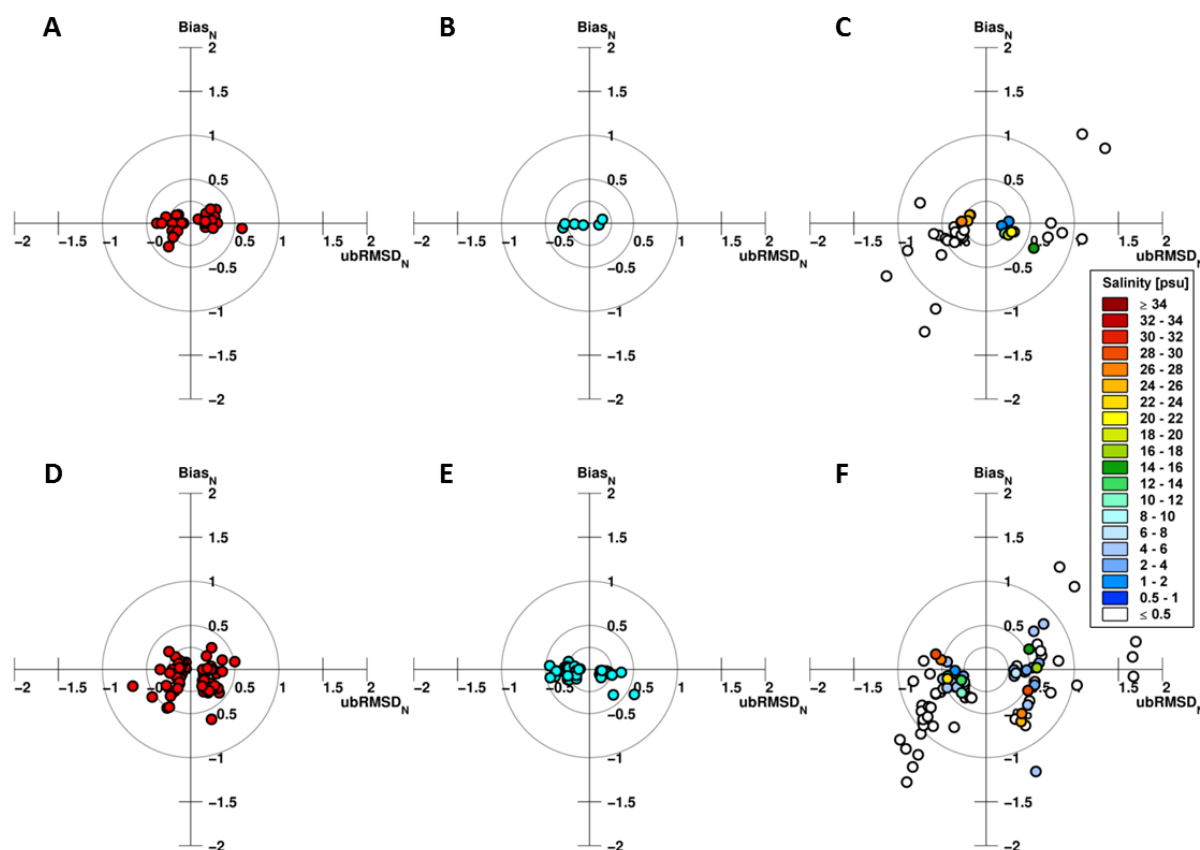
Figure 18 Target diagram summarizing how the model predictions compare to the time-series data during 1994-1997 for (**A**) water level, (**B**) tidal flow, (**C**) salinity, and during 2012 (MacWilliams and Bever 2013) for (**D**) water level, (**E**) tidal flow, (**F**) salinity. Symbols for salinity are colored based on the average predicted salinity at each station

Table 7 Summary of the number and percent of stations classified as very accurate, accurate, acceptable, and indicating poor agreement between predictions and observations based on the model skill and the target diagram accuracy thresholds shown in Table 1 for the 1994–1997 simulation period (this study) and for a simulation of calendar year 2012 (MacWilliams and Bever 2013)

		1994–1997				2012		
		Water level	Flow	Salinity	Current speed	Water Level	Flow	Salinity
Skill accuracy	Accurate	30 (88%)	5 (83%)	29 (76%)	14 (100%)	53 (75%)	29 (74%)	50 (66%)
	Acceptable	3 (9%)	1 (17%)	4 (11%)	0	12 (17%)	7 (18%)	14 (18%)
	Poor agreement	1 (3%)	0	5 (13%)	0	6 (8%)	3 (8%)	12 (16%)
Target accuracy	Very accurate	23 (67%)	4 (67%)	5 (13%)	3 (21%)	45 (63%)	21 (54%)	0
	Accurate	10 (30%)	2 (33%)	18 (48%)	9 (65%)	21 (30%)	17 (43%)	23 (30%)
	Acceptable	1 (3%)	0	9 (24%)	2 (14%)	5 (7%)	1 (3%)	40 (53%)
	Poor agreement	0	0	6 (15%)	0	0	0	13 (17%)

or "accurate" in 1994–1997 than in 2012, which indicates that the model predicted the data from the 1994–1997 stations slightly better than the data from the 2012 stations. Many of the stations for which data were available in 2012 but not in 1994–1997 are in locations that are difficult for the model to accurately predict, such as water level stations further upstream on the Sacramento River that are only weakly tidal, and salinity stations in the Central Delta that tend to have low average salinity and are also strongly affected by salinity from agricultural return flows. Including these stations slightly lowers the percentage of stations that fall within the accurate classification for 2012 compared to 1994–1997. However, these combined model skill and target diagram metrics demonstrate that the UnTRIM Bay–Delta Model accurately or acceptably predicted water level, flow, and salinity at the vast majority of stations during both simulation periods (Figure 18).

Setting a standard set of metrics with discrete thresholds for evaluating model performance is critical for comparing the performance of the simulations detailed in this study with other models that may be applied to the estuary. In addition, since the metrics applied here are not specific to the estuary, models of different regions could also be compared using the metrics and thresholds applied here. The thresholds used in this study to assess model accuracy (Table 1) were developed based on a review of previous studies. A more widespread application of this approach could be used to further refine the values used to set accuracy thresholds. Additional work could also be done to take into account how uncertainty in the observations affect the assessment of model accuracy.

X2 and the Low-Salinity Zone

One mechanism proposed for "fish–X2" relationships is that habitat area or volume varies with flow and X2 (Jassby et al. 1995; Kimmerer et al. 2009, 2013). In particular, there is a widespread belief that greater biological productivity and, potentially, fish population size results when the LSZ is adjacent to the broad shoals of Suisun Bay than when the LSZ is in the narrow confines of the Delta (Cloern et al. 1983; Jassby et al. 1995). Results of this modeling

study show that the area and volume of the LSZ do indeed increase as X2 moves from the Delta to Suisun Bay. However, with a further decrease in X2, the LSZ moves into the narrow confines of Carquinez Strait, and the area and volume of the LSZ decrease again (Figure 12). At extremely low X2 (high flow), LSZ area and volume increase again as X2 moves into into San Pablo Bay, but flow is rarely high enough for long enough to maintain this position for an extended period. Thus, the relationship between X2 and the area of the LSZ appears to be largely governed by the effective width of the estuary, with the largest LSZ area when X2 is in San Pablo or Suisun Bay, and the smallest LSZ area when X2 is in Carquinez Strait or in the junction region east of Mallard Island.

If the area and volume of the LSZ are important influences on biological productivity, unimodal responses of biological productivity to X2 over its usual range would be expected (~51 to 92 km based on mean for January through June). However, most of the relationships between X2 and abundance of several key fish species are monotonic (Jassby et al. 1995; Kimmerer et al. 2009, 2013), with little indication of a decrease in the slope of the relationship under high-flow conditions. Thus, our analysis and that by Kimmerer al. (2013) do not support the idea that the extent of the LSZ, as it is defined here between 0.5 and 6.0 psu, is strongly linked to biological productivity.

Since the accurate characterization of LSZ area in real time is inherently more difficult than real-time estimates of X2 based either on continuous salinity observations or X2 relationships, results from numerical models used to investigate the relationship between X2 and characteristics of the LSZ (or regions defined by other salinity ranges) can help to inform future regulations based on either outflow or X2. The area of the LSZ and the average depth of the LSZ shown on Figures 12 and 13 were calculated using the daily-averaged, depth-averaged salinity. However, the tidal excursion of the LSZ and its effect on the LSZ's extent must also be considered. The location of the 2-psu isohaline and the LSZ's areal extent both move over a wide region during each day because of tidal excursion (Figure 19). This is also evident in the

comparisons between observed and predicted salinity during spring and neap tides during 1994 (Figure 6 and Figure 7). When X2 is 75, the tidal excursion of the LSZ extends from eastern Suisun Bay to the western Delta. Honker Bay and Grizzly Bay remain within the LSZ during the entire day when X2 is 75, indicating a strong connectivity between the LSZ and the broad shoals of Suisun Bay throughout the tidal excursion.

X2 Autoregressive Equation

We developed a new autoregressive relationship for X2 as a function of Delta outflow (Equation 7) with a flow-varying weight between the flow term and the autoregressive term to predict X2. The free parameters were fit to a 3-year UnTRIM X2 time–series instead of X2 interpolated from continuous salinity observations, thereby eliminating errors associated with spatial interpolation and assumed top-to-bottom salinity differences. The Delta outflow used in the flow-X2 relationship developed here is the net Delta outflow from the Dayflow program (CDWR 1986, 2013). Delta outflow estimates produced by the Dayflow program contain substantial uncertainty, particularly during low Delta outflow conditions, because several terms in the volume balance are quite uncertain. Flow monitoring data collected since 1997 (Oltmann 1998) suggests that the actual daily-averaged Delta outflows can be very different from the reported Dayflow values. In particular, consumptive use in the Delta can only be estimated (either by QGCD in Dayflow or through the DICU estimates), and this can result in significant uncertainty in net Delta outflows. However, estimating outflow based on flow monitoring stations also results in significant uncertainty because it is very difficult to accurately derive a small net flow signal from a very large tidal flow signal that is many times greater than the net flow. Despite these uncertainties, we selected the Dayflow outflow estimates for this analysis because they have been used in previous autoregressive calculations and are readily available for other periods. Error in the UnTRIM X2 was not considered in the analysis because values of X2 interpolated from field observations have substantial uncertainty and therefore cannot be used to assess model error in X2. Instead, we

must rely on the accuracy of the model in representing near-bed salinity, as shown by the calibration. Therefore, the credible intervals reported for each day's output from the statistical model (Figure 14) are the range of X2 values that have a 95% probability of including the 'true' X2 value as determined using UnTRIM.

Both the constant-α equation and variable-α equation fits yield a significantly different exponent γ than reported by Monismith et al. (2002) (Table 6). The expected γ when salt flux is dominated by gravitational circulation is $-1/3$, while a γ of -1 can be expected when tidal stirring processes are dominant (MacCready 2007). The salt flux analysis in Gross et al. (2009) suggests that both these mechanisms are important in the estuary, and that the importance of gravitational circulation increases with flow as described by Monismith et al. (2002). With the improved X2 data set, our estimate for γ is -0.230 for the variable-α equation, and -0.202 for the constant-α equation. The difference between the two values indicates that neglecting the variability in time scale of response with flow biases the estimate of γ. The γ of -0.230 is substantially larger than the value of -0.141 estimated by Monismith et al. (2002), and also larger than the values of -0.17 and -0.20 found by Ralston et al. (2008) in subtidal salinity modeling of the estuary. Monismith et al. (2002) attribute the deviation from the $-1/3$ scaling to changes in stratification and salt flux mechanisms with flow, with gravitational circulation becoming stronger as flow and horizontal salinity gradients increase. Ralston et al. (2008) attribute the deviation from the $-1/3$ scaling primarily to along-axis variability in bathymetry. Our current analysis does not distinguish between these two explanations. However, we are planning further investigation of these physical mechanisms by analysis of the UnTRIM salt fluxes and subtidal salinity modeling.

The improvement in the variable-α equation is the use of a flow-dependent weighting between the flow term and the autoregressive term using a linear relationship between α and flow as suggested by Monismith et al. (2002). The weight of the autoregressive term decreases with flow to allow X2 to change more rapidly at high than at low Delta out-

A) Daily-average depth-averaged salinity
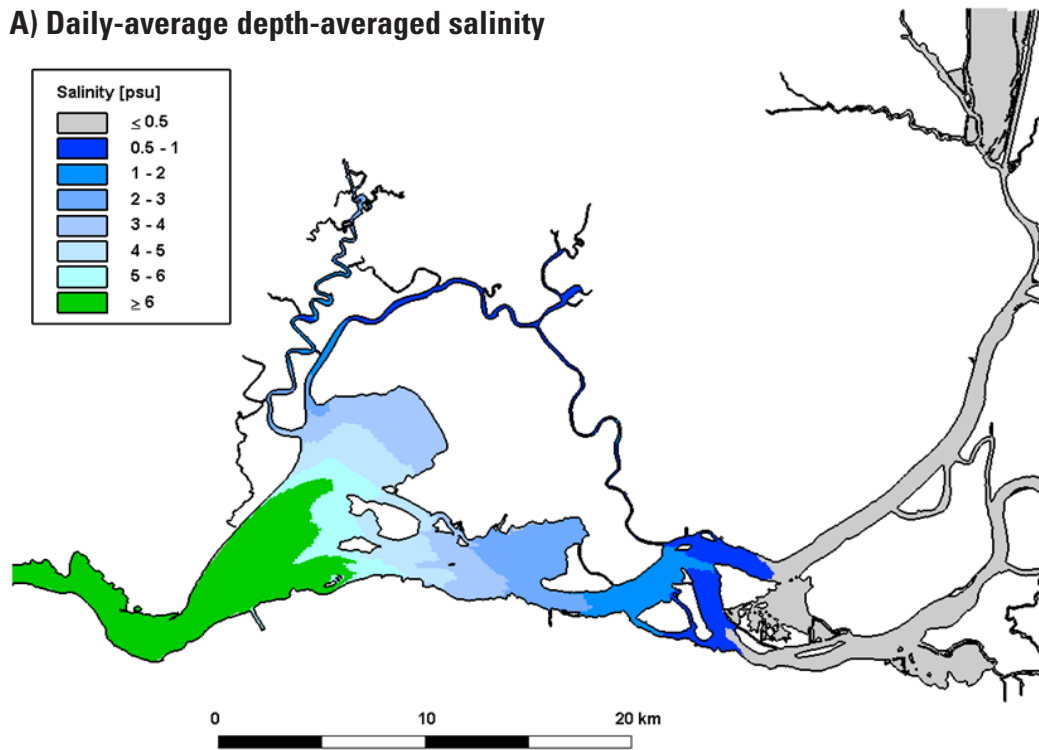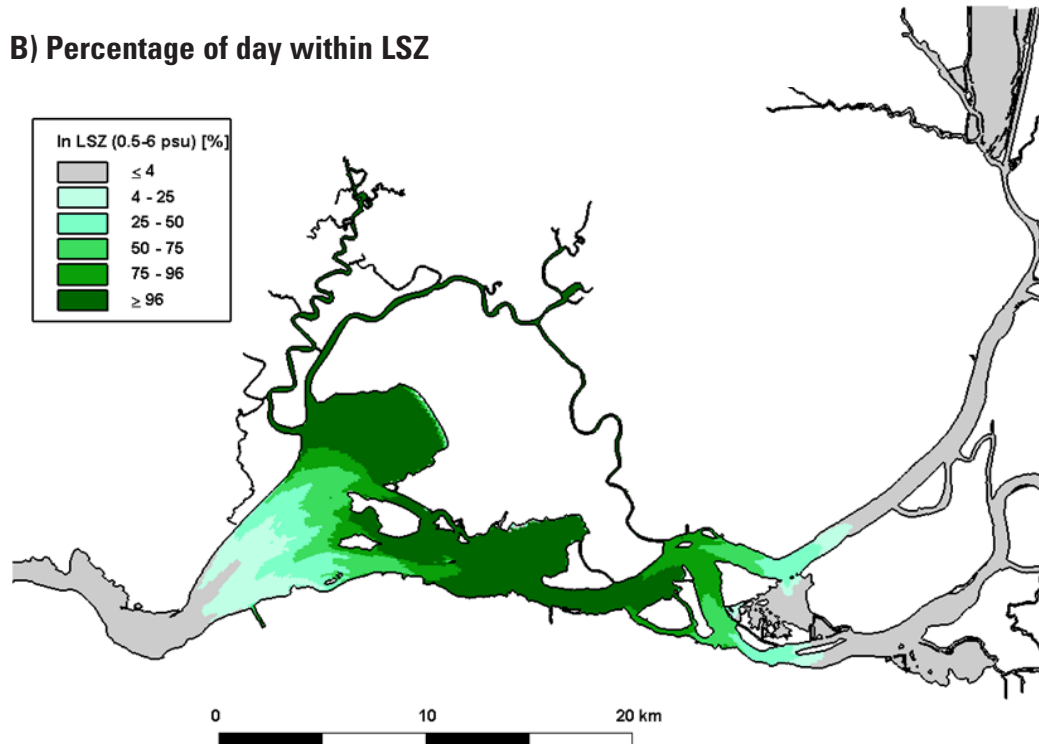

B) Percentage of day within LSZ

Figure 19 **(A)** Daily-averaged depth-averaged salinity between Carquinez Strait and the Western Delta on a day when X2 is approximately 75 km (see Figure 3 for X2 transect distances); **(B)** percent of the same day that the depth-averaged salinity is within the low-salinity zone (between 0.5 and 6 psu).

flow. The time scale of salinity adjustment associated with Equation 7 is related to α as $T=(1-\alpha)^{-1}$ (Monismith et al. 2002). In the constant-α equation, this yields a constant time scale of 9.52 days. In the variable-α equation $\alpha = \alpha_0 mQ + \alpha_0 b$. Because $\alpha_0 b$ is 1.00 (Table 5) the dependency of the time scale with flow is $T \sim Q^{-1}$. Lerczak et al. (2009) estimated time scales of adjustment to changes in freshwater outflow for two different representations of vertical mixing. In both cases the resulting time scale varies with both flow and X2 as $T \sim X2\,Q^{-1}$. Substituting Equation 8 for steady X2 into this relationship results in $T \sim Q^{-1.23}$. Therefore, the time scale associated with the variable-α equation has a somewhat different variability with flow than found in the analysis of Lerczak et al. (2009).

Monismith et al. (2002) reported that fitting X2 with a variable-α equation did not improve their regression to X2 estimated from continuous salinity observations. We believe this results from errors associated with estimating X2 from continuous salinity observations. In contrast the UnTRIM X2 dataset allowed improved accuracy for fitting a variable-α equation (Table 6). The variable-α equation predicts substantially different X2 values than previously published equations, particularly when UnTRIM X2 is high or low (Figure 15). The variable-α equation clearly captures more of the rapid variability in X2 during high outflows. Some differences between this equation and previous equations of the same form (Monismith et al. 2002; Gross et al. 2009) result from the different datasets used to provide daily X2 values. The Monismith et al. (2002) analysis used continuous observations of salinity from 1967–1990. Gross et al. (2009) used TRIM3D predictions of X2 for a data set that spanned two simulation periods: December 1996 to April 1998, and April 1994 to March 1995. Given that the salinity predictions presented in this paper indicate that the UnTRIM Bay–Delta model predicts observed synoptic and continuous observed salinity accurately and, in particular, more accurately than the TRIM3D model documented by Gross et al. (2009), we assert that the data set of X2 calculated using UnTRIM is superior to previous X2 data sets, even including those derived by interpolation of data from continuous surface salinity monitoring sta-

tions. Our assertion is supported by analysis of the error associated with estimating X2 from surface salinity observations (Table 4), which shows that the stratification assumption alone—only one of several assumptions needed to calculate X2 from surface salinity at fixed monitoring stations—results in a larger error than the error estimated for the UnTRIM predictions of X2.

Despite the significant improvements from previous autoregressive equations, the ~12 km credible intervals in this analysis (Figure 14) indicate that substantial uncertainty remains in X2 predictions. One major omission in all flow–X2 relationships is the lack of spring–neap variability which is clearly seen in the UnTRIM X2 values (Figure 11 and Figure 14). More generally, the X2 predicted by the variable-α equation often fails to match the high frequency (days to weeks) variability in UnTRIM X2. The effects of wind forcing and storm surge on salinity, which are not captured by the variable-α equation, also may be responsible for some of the variability in UnTRIM X2.

We have shown that the variable-α flow–X2 relationship introduced here is the most accurate approach for estimating X2 from Delta outflow alone (Table 6). This makes it highly useful for scientific studies in which accurate X2 estimates are needed and UnTRIM Bay–Delta model results are not available.

CONCLUSIONS

We used the 3-D UnTRIM San Francisco Bay–Delta model to simulate hydrodynamics and salinity in the San Francisco Estuary. A set of quantitative metrics were used to evaluate the accuracy of model predictions of water level, flow, and salinity. These metrics were shown to facilitate comparisons both between different models and between different periods simulated using the same model. We developed a quantitative approach to evaluate the accuracy of model predictions of X2 based on the observed near-bed salinity from the USGS water quality cruises. The predictions of X2 from the UnTRIM Bay-Delta model were shown to have less prediction error than X2 estimates which are derived from observed surface salinity.

We used the model to explore the relationships between X2 and the area and average depth of the low-salinity zone (LSZ). These results demonstrate that the relationships between X2 and the area and the average depth of the LSZ are not monotonic, and that the area and average depth of the LSZ are strongly influenced by geographic constrictions in Carquinez Strait and in the junction region of the Sacramento and San Joaquin rivers. The results from the UnTRIM Bay–Delta model simulations also have been used to refine relationships between physical habitat and species abundance (Kimmerer et al. 2013), and to drive a particle-tracking study to simulate the distribution of organisms using a range of vertical tidal migration behaviors (Kimmerer et al. 2014).

We used UnTRIM predictions of X2 to develop improved autoregressive equations that relate flow to X2. Two significant improvements over previous approaches were introduced. First, the weighting between the autoregressive term and the flow term was allowed to vary with Delta outflow, allowing the new variable-α equation to respond more strongly to changes in Delta outflow than previous regression relationships. Second, a Bayesian inference approach was used, which allows credible intervals to be generated for the X2 estimates. The two new flow–X2 relationships presented here are both shown to have lower Root Mean Square errors than three previously published relationships.

ACKNOWLEDGEMENTS

REFERENCES

Burau JR, Gartner JW, Stacey MT. 1998. Results for the hydrodynamic element of the 1994 Entrapment Zone Study in Suisun Bay. In: Kimmerer W, editor. Report of the 1994 Entrapment Zone Study. Interagency Ecological Program for the San Francisco Bay–Delta Estuary. p. 13–53.

Bever AJ, Friedrichs MAM, Friedrichs CT, Scully ME, Lanerolle LWJ. 2013. Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. J Geophys Res-Oceans 118:4924–4944. doi: *http://dx.doi.org/10.1002/jgrc.20331*

Bever AJ, Harris CK. 2014. Storm and fair-weather driven sediment-transport within Poverty Bay, New Zealand, evaluated using coupled numerical models. Cont Shelf Res 86:34–51. doi: *http://dx.doi.org/10.1016/j.csr.2013.07.012*

Bever AJ, MacWilliams ML. 2013. Simulating sediment transport processes in San Pablo Bay using coupled hydrodynamic, wave, and sediment transport models. Mar Geol 345:235–253. doi: *http://dx.doi.org/10.1016/j.margeo.2013.06.012*

Casulli V. 1999. A semi-implicit numerical method for non-hydrostatic free-surface flows on unstructured grid. In: Numerical modelling of hydrodynamic systems. Zaragoza, Spain: ESF Workshop. p. 175–193.

Casulli V. 2009. A high-resolution wetting and drying algorithm for free surface hydrodynamics. Int J Numer Meth Fl 60(4):391–408. doi: *http://dx.doi.org/10.1002/fld.1896*

Casulli V, Cattani E. 1994. Stability, accuracy and efficiency of a semi-implicit method for three-dimensional shallow water flow. Computers Mathematics Applications 27(4): 99-112.

Casulli V, Cheng RT. 1992. Semi-implicit finite difference methods for three-dimensional shallow water flow. Int J Numer Meth Fl 15(6):629–648. doi: *http://dx.doi.org/10.1002/fld.1650150602*

Casulli V, Stelling GS. 2011. Semi-implicit subgrid modelling of three-dimensional free-surface flows. Int J Numer Meth Fluids 67:441–449. doi: *http://dx.doi.org/10.1002/fld.2361*

Casulli V, Walters RA. 2000. An unstructured, three-dimensional model based on the shallow water equations. Int J Numer Meth Fl 32(3):331–348. doi: *http://dx.doi.org/10.1002/(SICI)1097-0363(20000215)32:3<331::AID-FLD941>3.0.CO;2-C*

Casulli V, Zanolli P. 2002. Semi-implicit numerical modelling of non-hydrostatic free-surface flows for environmental problems. Math Comput Model 36(9–10):1131–1149. doi: *http://dx.doi.org/10.1016/S0895-7177(02)00264-9*

Casulli V, Zanolli P. 2005. High resolution methods for multidimensional advection-diffusion problems in free-surface hydrodynamics. Ocean Model 10(1–2):137–151. doi: *http://dx.doi.org/10.1016/j.ocemod.2004.06.007*

Casulli V, Zanolli P. 2007. Comparing analytical and numerical solution of nonlinear two and three-dimensional hydrostatic flows. Int J Numer Meth Fl 53(6):1049–1062. doi: *http://dx.doi.org/10.1002/fld.1358*

[CDEC] California Data Exchange Center. 2013. Daily X2 calculation [Internet]. [accessed 1 Dec 2013]. Available from: *http://cdec.water.ca.gov/cgi-progs/stationInfo?station_id=CX2*

[CDWR] California Department of Water Resources. 1986. DAYFLOW program documentation and data summary user's guide. Sacramento (CA): California Department of Water Resources. Available from: *http://www.water.ca.gov/dayflow/docs/DAYFLOW_1986.pdf*

[CDWR] California Department of Water Resources. 1995. Estimation of Delta island diversions and return flows. Sacramento (CA): Modeling Support Branch, Division of Planning, California Department of Water Resources. Available from: *http://www.calwater.ca.gov/Admin_Record/C-032892.pdf*

[CDWR] California Department of Water Resources. 2008. Delta Simulation Model II documentation, Available from: *http://baydeltaoffice.water.ca.gov/modeling/deltamodeling/models/dsm2/dsm2.cfm*

[CDWR] California Department of Water Resources. 2013. Dayflow, an estimate of daily average Delta outflow [Internet]. [accessed 7 Nov 2013]. Available from: *http://www.water.ca.gov/dayflow/*

[CSUMB] California State University, Monterey Bay. 2005. Data library, Central California data, Seafloor Mapping Lab; [accessed 19 Jan 2015]. Available from: *http://seafloor.csumb.edu/SFMLwebDATA_c.htm#SFB*

Cheng RT, Casulli V, Gartner JW. 1993. Tidal Residual Intertidal Mudflat (TRIM) model and its applications to San Francisco Bay, California. Estuar Coast Shelf Sci 36(3):235–280. doi: *http://dx.doi.org/10.1006/ecss.1993.1016*

Cheng RT, Gartner JW, Smith RE. 1997. Bottom Boundary Layer in South San Francisco Bay, California. J Coastal Res. 25(S1):49–62.

Cheng P, Wilson RE, Chant RJ, Fugate DC, Flood RD. 2009. Modeling influence of stratification on lateral circulation in a stratified estuary. J Phys Oceanogr 39:2324–2337. doi: *http://dx.doi.org/10.1175/2009JPO4157.1*

Cloern JE. 1984. Temporal dynamics and ecological significance of salinity stratification in an estuary (South San Francisco Bay, USA). Oceanol Acta 7:137–141. Available from: *http://archimer.ifremer.fr/doc/00113/22416/*

Cloern J, Alpine A, Cole B, Wong R, Arthur J, Ball M. 1983. River discharge controls phytoplankton dynamics in the northern San Francisco Bay estuary. Estuar Coast Shelf Sci 16(4):415–429. doi: *http://dx.doi.org/10.1016/0272-7714(83)90103-8*

Deleersnijder E, Beckers JM, Campin JM, El Mohajir M, Fichefet T, Luyten P. 1997. Some mathematical problems associated with the development and use of marine models. In: Diaz JI, editor. The mathematics of models for climatology and environment. Vol. 48. Berlin, Heidelberg: Springer–Verlag. p. 39–86. doi: *http://dx.doi.org/10.1007/978-3-642-60603-8_2*

Dever EP, Lentz SJ. 1994. Heat and salt balances over the northern California shelf in winter and spring. J Geophys Res 99(C8):16001–16017. doi: *http://dx.doi.org/10.1029/94JC01228*

Elias EPL, Hansen JE. 2013. Understanding processes controlling sediment transports at the mouth of a highly energetic inlet system. Mar Geol 345:207–220. doi: *http://dx.doi.org/10.1016/j.margeo.2012.07.003*

Feyrer F, Nobriga M, Sommer T. 2007. Multi-decadal trends for three declining fish species: Habitat patterns and mechanisms in the San Francisco Estuary, California, USA. Can J Fish Aquat Sci 64(4):723–734. doi: *http://dx.doi.org/10.1139/F07-048*

Foxgrover AC, Jaffe BE, Hovis GT, Martin CA, Hubbard JR, Samant MR, Sullivan SM. 2007. 2005 Hydrographic Survey of South San Francisco Bay, California, U.S. Geological Survey, Open File Report 2007–1169.

Fram JP, Martin MA, Stacey MT. 2007. Dispersive fluxes between the coastal ocean and a semienclosed estuarine basin. J Phys Oceanogr 37(6):1645–1660. Available from: *http://dx.doi.org/10.1175/JPO3078.1*

Gardner JV, Mayer MA, Hughes–Clarke J. 1998. Cruise report R/V Inland Surveyor cruise IS-98. The bathymetry of Lake Tahoe, California-Nevada. U.S. Geological Survey Open-File Report 98-509.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. Bayesian data analysis, 2nd ed. Boca Raton (FL): CRC Press.

Gross ES, MacWilliams ML, Kimmerer W. 2006. Simulating periodic stratification in San Francisco Bay. In: Spaulding M, editor. Proceedings of the Estuarine and Coastal Modeling Conference, ASCE.

Gross ES, MacWilliams ML, Kimmerer WJ, 2009. Three-dimensional modeling of tidal hydrodynamics in the San Francisco Estuary. San Franc Estuary Watershed Sci [Internet]. [cited 13 Dec 2013];7(2). Available from: *https://escholarship.org/uc/item/9rv243mg* doi: *http://dx.doi.org/10.15447/sfews.2009v7iss2art2*

Haidvogel DB, Arango H, Budgell WP, Coruelle BD, Curshitser E, Lorenzo ED, Fennel K, Geyer WR, Hermann AJ, Lanerolle L, et al. 2008. Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. J Computat Phys 227(7):3595–3624. doi: *http://dx.doi.org/10.1016/j.jcp.2007.06.016*

Hofmann EE, Cahill B, Fennel K, Friedrichs MAM, Hyde K, Lee C, Mannino A, Najjar RG, O'Reilly JE, Wilkin J, et al. 2011. Modeling the dynamics of continental shelf carbon. Annu Rev Mar Sci 3:93–122. doi: *http://dx.doi.org/10.1146/annurev-marine-120709-142740*

Jassby AD, Kimmerer WJ, Monismith SG, Armor C, Cloern JE, Powell TM, Schubel JR, Vendlinski TJ. 1995. Isohaline position as a habitat indicator for estuarine populations. Ecol Appl 5(1):272–289. doi: *http://dx.doi.org/10.2307/1942069*

Jolliff JK, Kindle JC, Shulman I, Penta B, Friedrichs MAM, Helber R, Arnone RA. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. J Mar Syst 76(1–2):64–82. doi: *http://dx.doi.org/10.1016/j.jmarsys.2008.05.014*

Kantha LH, Clayson CA. 1994. An improved mixed layer model for geophysical applications. J Geophys Res 99(C12):25235–25266. doi: *http://dx.doi.org/10.1029/94JC02257*

Kimmerer WJ. 2002. Physical, biological, and management responses to variable freshwater flow into the San Francisco Estuary. Estuaries 25(6B):1275–1290. doi: *http://dx.doi.org/10.1007/BF02692224*

Kimmerer WJ, Bennett WA, Burau JR. 2002. Persistence of tidally-oriented vertical migration by zooplankton in a temperate estuary. Estuaries 25(3):359–371. doi: *http://dx.doi.org/10.1007/BF02695979*

Kimmerer WJ, Burau JR, Bennett WA. 1998. Tidally-oriented vertical migration and position maintenance of zooplankton in a temperate estuary. Limnol Oceanogr 43(7):1697–1709. doi: *http://dx.doi.org/10.4319/lo.1998.43.7.1697*

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? Estuar Coast. 32(2):375–389. doi: http://dx.doi.org/10.1007/s12237-008-9124-x

Kimmerer WJ, Gross ES, MacWilliams ML. 2014. Tidal migration and retention of estuarine zooplankton investigated using a particle-tracking model. Limnol Oceanogr 59(3):901–916. doi: http://dx.doi.org/10.4319/lo.2014.59.3.0901

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 [Internet]. [cited 13 Dec 2013];11(4). Available from: https://escholarship.org/uc/item/3pz7x1x8 doi: http://dx.doi.org/10.15447/sfews.2013v11iss4art1

Large W, Pond S. 1981. Open ocean momentum flux measurements in moderate to strong winds. J Phys Oceanogr 11(3):324–336. Available from: http://dx.doi.org/10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2

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 39(4):915–933. Available from: http://dx.doi.org/10.1175/2008JPO3933.1

Li M, Zhong L, Boicort WC. 2005. Simulations of Chesapeake Bay estuary: sensitivity to turbulence mixing parameterizations and comparison with observations. J Geophys Res 110(C12). doi: http://dx.doi.org/10.1029/2004JC002585

Lippert C, Sellerhoff F. 2006. Efficient generation of unstructured orthogonal grids. Proceedings of the 7th International Conference on Hydroscience and Engineering; 2006 Sep 10–13; Philadelphia. Available from: https://idea.library.drexel.edu/islandora/object/idea%3A1472

Liu Y, MacCready P, Hickey BM, Dever EP, Kosro PM, Banas NS. 2009. Evaluation of a coastal ocean circulation model for the Columbia River plume in summer 2004. J Geophys Res 114(C2). doi: http://dx.doi.org/10.1029/2008JC004929

Ma G, Shi F, Liu S, Qi D. 2011. Hydrodynamic modeling of Chiangjian Estuary: Model skill assessment and large-scale structure impacts. Appl Ocean Res 33(1):69–78. doi: http://dx.doi.org/doi:10.1016/j.apor.2010.10.004

MacCready P. 2007. Estuarine adjustment. J Phys Oceanogr 27:2133–2145. doi: http://dx.doi.org/10.1175/JPO3082.1

MacWilliams ML. 2004. Three-dimensional hydrodynamic simulation of river channels and floodplains [dissertation]. [Stanford (CA)]: Stanford University.

MacWilliams ML, Bever AJ. 2013. 3-D hydrodynamic simulations for Delta Smelt hatching distribution studies, Federal Science Task Force Fisheries Investigations. Prepared for Science Division, Bay–Delta Office, U.S. Bureau of Reclamation, U.S. Department of Interior. 235 p.

MacWilliams ML, Bever AJ, Gross ES. 2012a. Three-dimensional sediment transport modeling for San Francisco Bay RDMMP, San Francisco Bay Regional dredged material management plan. Prepared for U.S. Army Corps of Engineers, San Francisco District. 190 p.

MacWilliams ML, Kilham NE, Bever AJ. 2012b. South San Francisco Bay long wave modeling report, South San Francisco Bay shoreline study. Prepared for U.S. Army Corps of Engineers, San Francisco District. 226 p.

MacWilliams ML, Cheng RT. 2006. Three-dimensional hydrodynamic modeling of San Pablo Bay on an unstructured grid. Proceedings of the 7th International Conference on Hydroscience and Engineering; 2006 Sep 10–13; Philadelphia. Available from: https://idea.library.drexel.edu/islandora/object/idea%3A1438

MacWilliams ML, Gross ES. 2007. UnTRIM San Francisco Bay–Delta Model calibration report, Delta Risk Management Study. Prepared for the California Department of Water Resources. 101 p.

MacWilliams ML, Gross ES. 2010. UnTRIM San Francisco Bay–Delta Model sea level rise scenario modeling report, Bay–Delta Conservation Plan. Prepared for Science Applications International Corporation and California Department of Water Resources. 562 p.

MacWilliams ML, Gross ES. 2013. Hydrodynamic simulation of circulation and residence time in Clifton Court Forebay. San Franc Estuary Watershed Sci [Internet]. [cited 13 Dec 2013];11(2). Available from: *http://www.escholarship.org/uc/ item/4q82g2bz* doi: *http://dx.doi.org/10.15447/ sfews.2013v11iss2art1*

MacWilliams ML, Salcedo FG, Gross ES. 2008. San Francisco Bay–Delta UnTRIM Model calibration report, POD 3-D Particle Tracking Modeling Study. Prepared for California Department of Water Resources. 344 p.

MacWilliams ML, Salcedo FG, Gross ES. 2009. San Francisco Bay–Delta UnTRIM Model calibration report, Sacramento and Stockton Deep Water Ship Channel 3-D Hydrodynamic and Salinity Modeling Study. Prepared for U.S. Army Corps of Engineers, San Francisco District. 574 p.

Marvin Jung & Associates, Inc. 2000. Revision of Representative Delta Island Return Flow Quality for DSM2 and DICU model runs. Prepared for the CALFED Ad-Hoc Workgroup to Simulate Historical Water Quality Conditions in the Delta by Marvin Jung & Associates, Inc. Available from: *http:// www.water.ca.gov/waterquality/drinkingwater/docs/ modeling%20reports/revision_of_representative_delta_ island_return_flow_quality_for_dsm2_and_dicu_ model_runs._december_2000.pdf*

Monismith SG, Burau JR, Stacey M. 1996. Stratification dynamics and gravitational circulation in northern San Francisco Bay. In: Hollibaugh JT, editor. San Francisco Bay: the ecosystem. San Francisco (CA): Pacific Division, AAAS. p. 123–153.

Monismith SG, Kimmerer W, Burau JB, Stacey MT. 2002. Structure and flow-induced variability of the subtidal salinity field in Northern San Francisco Bay. J Phys Oceanogr 32(11):3003–3019. doi: *http:// dx.doi.org/10.1175/1520-0485(2002)032<3003:SAF IVO>2.0.CO;2*

[NOAA] National Oceanic and Atmospheric Administration. 1998. San Francisco Bay, CA (P090) bathymetric digital elevation model (30-meter resolution) derived from source hydrographic survey soundings collected by NOAA. [cited 19 Jan 2015]. Available from: *http://estuarinebathymetry.noaa.gov/ bathy_htmls/P090.html*

Nobriga M, Sommer T, Feyrer F, Fleming K. 2008. Long-term trends in summertime habitat suitability for delta smelt, *Hypomesus transpacificus*. San Franc Estuary Watershed Sci [Internet]. [cited 13 Dec 2013];6(1). Available from: *https://escholarship.org/ uc/item/5xd3q8tx* doi: *http://dx.doi.org/10.15447/ sfews.2008v6iss1art1*

Oltmann RN. 1998. Indirect measurement of delta outflow using ultrasonic velocity meters and comparison with mass-balance calculated outflow. Interagency Ecological Program for the Sacramento–San Joaquin Estuary Newsletter [Internet] 11(1). Available from: *http://www.water. ca.gov/iep/newsletters/1998/winter/Indirect%20 Measurement%20of%20Delta%20Outflow%20 Using%20Ultrasonic%20Velocity%20Meters.pdf*

Ralston DK, Geyer WR, Lerczak JA. 2008. Subtidal salinity and velocity in the Hudson River estuary: observations and modeling. J Phys Oceanogr 38(4):753–770. doi: *http://dx.doi. org/10.1175/2007JPO3808.1*

Ralston DK, Geyer WR, Lerczak JA. 2010. Structure, variability, and salt flux in a strongly forced salt wedge estuary. J Geophys Res 115(C06005). doi: *http://dx.doi.org/10.1029/2009JC005806*

[RMA] Resource Management Associates. 2005. Flooded islands pre-feasibility study: RMA Delta Model calibration report. Prepared for the California Department of Water Resources for submittal to California Bay–Delta Authority. 158 p. Available from: *http://www.water.ca.gov/frankstract/ docs/%288%29RMA-Calibration%20Report.pdf*

[SCCOOS] Southern California Coastal Ocean Observing System. 2012. SIO manual shore stations program, SE Farallon Island shore station. Available from: *http://www.sccoos.org*

Spiegelhalter D, Thomas A, Best N. 2003. WinBUGS version 1.4. Bayesian inference using Gibbs sampling. Cambridge (UK): Medical Research Council, Biostatistics Unit, Institute for Public Health.

Stacey MT, Burau JR, Monismith SG, 2001. Creation of residual flows in a partially stratified estuary. J Geophys Res 106(C8):17013–17037. doi: *http://dx.doi.org/10.1029/2000JC000576*

[SWRCB] State Water Resources Control Board, California Environmental Protection Agency. 2000. Revised Water Right Decision 1641, In the Matter of: Implementation of Water Quality Objectives for the San Francisco Bay/Sacramento–San Joaquin Delta Estuary; A Petition to Change Points of Diversion of the Central Valley Project and the State Water Project in the Southern Delta; and A Petition to Change places of Use of the Central Valley Project, December 29, 1999; Revised in Accordance with Order WR 2000-02.

Stow AC, Jolliff J, McGillicuddy Jr DJ, Doney SC, Allen JI, Friedrichs MAM, Rose KA, Wallhead P. 2009. Skill assessment for coupled biological/physical models of marine systems. J Mar Syst 76(1–2):4–15. doi: *http://dx.doi.org/10.1016/j.jmarsys.2008.03.011*

Towill, Inc. 2009. San Francisco to Stockton DWSC 3D models. Prepared for United States Army Corps of Engineers, San Francisco District.

Umlauf L, Burchard H. 2003. A generic length-scale equation for geophysical turbulence models. J Mar Res 61(2):235–265. doi: *http://dx.doi.org/10.1357/002224003322005087*

[USGS] United States Geological Survey. 2007. San Francisco Bay bathymetry. Available from: *http://walrus.wr.usgs.gov/pacmaps/sf_index.html*

[USGS] United States Geological Survey. 2013. USGS Water Quality of San Francisco Bay. [accessed 2015 Jan 19]. Available from: *http://sfbay.wr.usgs.gov/access/wqdata*

van Leer B. 1979. Towards the ultimate conservative difference scheme. V. A second order sequel to Godunov's method. J Comput Phys 32(1):101–136. doi: *http://dx.doi.org/10.1016/0021-9991(79)90145-1*

Warner JC, Geyer WR, Lerczak JA. 2005a. Numerical modeling of an estuary: A comprehensive skill assessment, J Geophys Res 110(C05001). doi: *http://dx.doi.org/10.1029/2004JC002691*.

Warner JC, Sherwood CS, Arango HG, Signell RP. 2005b. Performance of four turbulence closure models implemented using a generic length scale method. Ocean Model. 8(1–2):81–113. doi: *http://dx.doi.org/10.1016/j.ocemod.2003.12.003*

Willmott CJ. 1981. On the validation of models. Phys Geogr 2:184–194. Available from: *http://climate.geog.udel.edu/~climate/publication_html/Pdf/W_Phys_Geog_1981.pdf*

Zorndt A, Schlurmann T, Grabemann I. 2012. The influence of extreme events on hydrodynamics and salinities in the Weser Estuary in the context of climate impact research. Coast Eng Proc 1(33):currents.50. doi: *http://dx.doi.org/10.9753/icce.v33.currents.50*

NOTES

Smith RE, Foxgrover A, Smith PE. 2003. Suisun Bay and Delta bathymetry: production of a 10-meter grid. Poster presented at: Second Biennial CALFED Science Conference: Advances in Science and Restoration in the Bay, Delta and Watershed; Sacramento, CA. Available from: *http://sfbay.wr.usgs.gov/sediment/delta/Delta_bathy_CALFED.pdf*