Virtual Special Issue
Coastal ocean modelling

Numerical modeling of circulation in high-energy estuaries: A Columbia River estuary benchmark

Tuomas Kärnä [a],*, António M. Baptista [a], Jesse E. Lopez [a], Paul J. Turner [a], Craig McNeil [b], Thomas B. Sanford [b]

[a] NSF Science and Technology Center for Coastal Margin Observation & Prediction, Oregon Health & Science University, Portland, OR, USA
[b] NSF Science and Technology Center for Coastal Margin Observation & Prediction, University of Washington, Seattle, WA, USA

ABSTRACT

Numerical modeling of three-dimensional estuarine circulation is often challenging due to complex flow features and strong density gradients. In this paper the skill of a specific model is assessed against a high-resolution data set, obtained in a river-dominated mesotidal estuary with autonomous underwater vehicles and a shipborne winched profiler. The measurements provide a detailed view of the salt wedge dynamics of the Columbia River estuary. Model skill is examined under contrasting forcing conditions, covering spring freshet and autumn low flow conditions, as well as spring and neap tides. The data set provides a rigorous benchmark for numerical circulation models. This benchmark is used herein to evaluate an unstructured grid circulation model, based on linear finite element and finite volume formulations. Advection of momentum is treated with an Eulerian–Lagrangian scheme. After the model's sensitivity to grid resolution and time step is examined, a detailed skill assessment is provided for the best model configuration. The simulations reproduce the timing and tidal asymmetry of salinity intrusion. Sharp density gradients, however, tend to be smoothed out affecting vertical mixing and gravitational circulation. We show that gravitational salt transport is underestimated in the model, but is partially compensated through tidal effects. The discrepancy becomes most pronounced when the stratification is strongest, i.e., under high river discharge and neap tide conditions.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Estuaries serve as active buffers between rivers and the continental shelf, transforming riverine fluxes of carbon and other nutrients before they reach the sea. Detailed understanding of the physical and biogeochemical processes that take place in estuaries is crucial for assessing their ecosystem function, and their response to climate change and other anthropogenic pressures. To this end, numerical modeling, paired with high-resolution measurements, offer an excellent opportunity to advance estuarine science.

However, numerical modeling of high-energy estuaries remains challenging due to strong tidal currents, sharp salinity gradients, and a wide range of relevant length and time scales. Simulating strong currents over a complex topography imposes constraints on spatial and temporal resolution. Sharp density gradients, an essential feature of strongly stratified estuaries, can only be captured using accurate numerical transport schemes with low

numerical mixing. In addition, the complex interplay of momentum and density fields, including turbulent processes, must be accurately represented. To capture non-linear interactions across scales, it is increasingly recognized that computational domains should extend from the upstream boundary of tidal rivers to the shelf break (or beyond), adding to the complexity and computational cost of the problem.

While many benchmarks exist for circulation models, there are few realistic test cases for high-energy estuaries based on observational data. This paper introduces a rigorous benchmark for the Columbia River estuary, located on the coast of the Northeast Pacific Ocean (Fig. 1). We rely on observational data from the SATURN collaboratory (Science And Technology University Research Network, Baptista et al., 1998) that includes an extensive river-to-shelf observation network and a skill-assessed modeling system. Multiple endurance stations measure both physical and biochemical variables in the river, estuary and continental shelf sea. Data coverage is further extended with seasonal glider and autonomous underwater vehicle (AUV) operations and specific measurement campaigns.

The Columbia River estuary is a mesotidal system subject to intense summer upwelling, but characterized by high river flow. The contemporary annual mean discharge is $5500 \, \text{m}^3 \, \text{s}^{-1}$ and spring freshet typically exceeds $10\,000 \, \text{m}^3 \, \text{s}^{-1}$ (Chawla et al., 2008). The maximum daily tidal range varies from less than 2 m during neaps to 3.6 m in springs, and tidal currents can exceed $3 \, \text{m} \, \text{s}^{-1}$ near the mouth.

The rapid flushing time, high stratification and relatively strong tidal influence make the Columbia River estuary an exceptional system (Geyer, 2010). Most estuaries in the same latitudes tend to have smaller freshwater input, whereas large rivers closer to the equator have weaker tides. Also, few large-discharge estuaries are exposed to the strong coastal upwelling characteristic of an Eastern Boundary Current. As such, the estuary constitutes a challenging benchmark for both studying estuarine processes and testing numerical models.

The circulation in the estuary is largely determined by three primary forcings: river discharge, tides and coastal winds (Chawla et al., 2008). River discharge is modulated by the operation of the Federal Columbia River Power System, a large and economically important system with primarily seasonal storage capacity. In spite of that modulation, seasonal variability remains strong (Fig. 2a). The discharge is highest during the spring freshet, which occurs typically in May–June in response to snow melt in the Eastern part of the watershed. Flows are maintained artificially low during most of the summer and fall, to enable moderate flows for power production during winter months.

Tides are mixed semidiurnal, with the two major tidal constituents being M_2 (amplitude $\approx 0.95 \, \text{m}$) and K_1 ($\approx 0.41 \, \text{m}$) (Chawla et al., 2008). In addition to driving the surface elevations, the tides have significant impact on mixing in the lower estuary. All tidal constituents exhibit temporal variability in response to strong nonlinear interactions with the river discharge (Jay and Flinchem, 1997). Winds in the continental shelf determine coastal up- or downwelling (Hickey and Banas, 2003), thus influencing the salt content and density structure of the waters entering the estuary during flood tides. Compared to the effect of river discharge, tides and coastal upwelling, the direct impact of local winds is less important to the estuarine circulation (Chawla et al., 2008; Elias et al., 2012).

The estuary is moderately to highly stratified during low flow conditions, shifting towards a salt wedge system as flows increase (Hansen and Rattray, 1966; Hughes and Rattray, 1980; Geyer and MacCready, 2014). Under high flows the effect of the freshwater discharge dominates, enabling the lower estuary to remain highly stratified across all tidal conditions (Jay and Smith, 1990).

By contrast, at low flows the salinity structure is tidally dependent (Jay and Smith, 1990; Jay and Smith, 1990). During moderate and spring tides the estuary is weakly stratified. Vertical exchange between the saline and freshwater layers is dominated by bottom friction-induced turbulence, which extends upwards into most of the water column due to the weak stratification. Moving towards neap tides, tidal energy decreases, resulting in reduced mixing and increased stratification. Under these conditions, the bottom layer enters the lower estuary almost without dilution. Internal oscillations induced by shear instabilities at the interface dominate vertical exchange between the layers (Jay and Smith, 1990; Kay and Jay, 2003).

Two deep channels in an otherwise shallow estuary (Fig. 1) are responsible for most estuarine transport. These channels have markedly different circulation characteristics: the North Channel is flood-dominant, while the South Channel is ebb-dominant and a preferential conduit for the freshwater discharge (Jay and Smith, 1990; Chawla et al., 2008). In the North Channel, residual inflow (as measured by the dimensionless inflow number of Chawla et al., 2008) is strong, and occurs almost continuously, being suppressed only during spring tides and peak discharges. By contrast, in the South Channel river discharge effectively counteracts gravitational circulation resulting in weaker inflow; for flows above $8000 \, \text{m}^3 \, \text{s}^{-1}$ the inflow is entirely suppressed (Chawla et al., 2008). In the South Channel residual inflow is also strongly modulated by tides. It is strongest during neap tides, whereas it is nearly absent during spring tides even under low flow conditions.

Upstream salt transport is driven by gravitational circulation and tidal mechanisms. Hughes and Rattray (1980) studied three cross-sections in the lower estuary during both high and low flow conditions and found that over half of the up-estuary salt flux is due to tidal pumping with the remainder being attributed to gravitational circulation. This finding is supported by subsequent studies as well (e.g., Jay and Smith, 1990; Chawla et al., 2008).

Early numerical modeling of the estuarine circulation in the Columbia River estuary was conducted by Hamilton (1990), who used a network of laterally averaged models to study salinity intrusion under different river discharge conditions. During the 1990s, various depth-averaged (ADCIRC, Luettich et al., 1992; WET2D, Beck and Baptista, 1997) and three-dimensional (QUODDY, Lynch et al., 1996; POM, Blumberg and Mellor, 1987) circulation models were implemented for the Columbia River, as potential computational engines for a pioneering observation and prediction system (Baptista et al., 1998). Each of these models failed to represent the estuarine dynamics in a satisfactory manner, but lessons learned led to the development of ELCIRC (Zhang et al., 2004), a three-dimensional Eulerian–Lagrangian finite volume model inspired by UNTRIM (Casulli and Walters, 2000) and designed specifically to meet the high-energy/high-stratification challenges posed by the Columbia River estuary.

ELCIRC permitted the first meaningful fully three-dimensional simulations of the entire river-estuary-plume-shelf system (Baptista et al., 2005), and offered a consistently reliable computational engine to conduct both daily forecasts and extended multi-year hindcast simulations. However, ELCIRC retained a number of
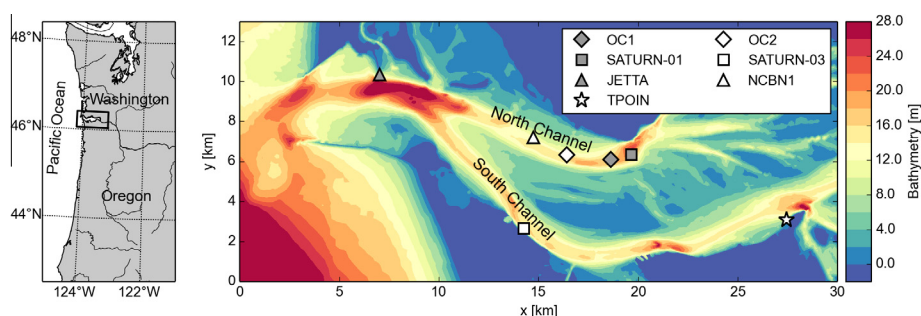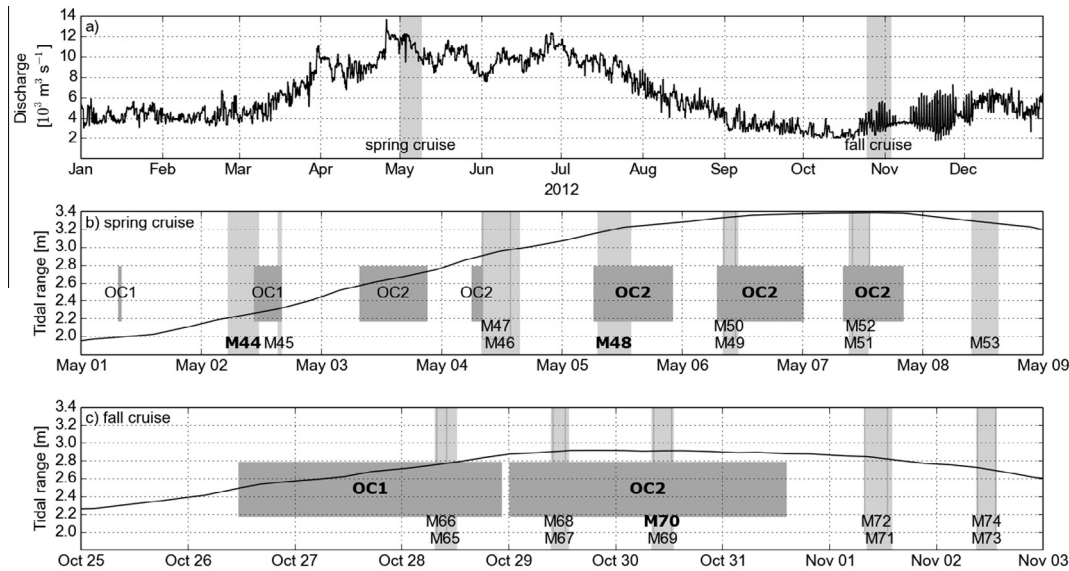

Fig. 1. Bathymetry of the lower estuary with station locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. Summary of collected data sets. (a) Columbia River discharge for 2012 in Bonneville dam (235 km upstream from the mouth). Tidal range at TPOIN during the (b) spring and (c) fall cruise. Light gray shading in (b) and (c) mark the time span of AUV missions, while darker rectangles mark the winched profiler operations at OC1 and OC2 locations. Data sets labeled with bold face are analyzed in detail.

undesirable constraints including the requirement for orthogonal unstructured grid in the horizontal and a z coordinate system in the vertical. The model also tended to underrepresent salinity intrusion and was generally too diffusive. The desire to alleviate these shortcomings lead to the development of the finite element model SELFE (Zhang and Baptista, 2008), which yielded better accuracy while retaining similar computational efficiency. SELFE performed better within the estuary and on the continental shelf (Zhang and Baptista, 2008; Burla et al., 2010), and currently serves as the core of the SATURN modeling system.

In the literature other circulation models have been used for the Columbia system as well. In MacCready et al. (2009) the estuary-plume-shelf system was modeled with ROMS (Haidvogel et al., 2000) for the summer 2004, focusing on the influence of fresh water input on vertical mixing both in the estuary and the plume regions. The model was validated against velocity, temperature and salinity data on the continental shelf. A more elaborate skill analysis of the model is presented in Liu et al. (2009), using an extensive set of measurements on the shelf and in the estuary. Elias et al. (2012) calibrated a coupled circulation and wave model (Lesser et al., 2004, Delft3D-SWAN,) in the Columbia River estuary using observations at the mouth.

This paper presents a detailed model skill assessment, based on a high-resolution observational data set obtained in the Columbia River estuary. The data were collected in the North Channel during two cruises in 2012, in May (henceforth the spring cruise) and October (the fall cruise), corresponding to high and low flow regimes, respectively (Fig. 2a). The instrumentation included two AUVs and the winched profiler, a ship operated, winched multi-instrument profiling platform. We present an overview and specific examples of the data set, to illustrate the dynamics of salinity intrusion in the Columbia River estuary.

The simulations are carried out with the SELFE model. First a calibration of the model grid resolution and time step is presented. SELFE does not have a strict CFL (Courant–Friedrichs–Lewy) stability constraint on time step due to the Eulerian–Lagrangian formulation and semi-implicit treatment of the free surface. However, we demonstrate that the time step must be chosen carefully to capture the salt dynamics: reasonable salinity intrusion was obtained only with sufficient spatial resolution matched with a compatible time step. The model's sensitivity to bottom friction is

also demonstrated. In the latter part of the paper, the skill of the best model setup is contrasted with the AUV and winched profiler observations in detail, focusing on salinity and current structure.

The paper is organized as follows. Section 2 introduces the observational data sets, followed by an outline of the numerical model. Model-observation comparisons are presented in Section 3. Sensitivity to grid resolution and time step, as well as bottom friction, is demonstrated in Sections 3.1 and 3.2. Overall skill of the best model configuration is presented in 3.3. More detailed comparison against selected AUV and winched profiler observations is presented in Sections 3.4 and 3.5. Section 4 summarizes the observed salt dynamics and the model skill. Conclusions are drawn in Section 5.

2. Methods

2.1. Observations

Two cruises were carried out in the North Channel of the estuary in May (spring cruise) and October (fall cruise) 2012, corresponding to typical high and low river discharge conditions (Fig. 2a). During the cruises high-resolution measurements were collected using both the shipborne winched profiler and AUVs, focusing on salinity intrusion and suspended particulate matter dynamics. The ship was anchored at two locations, OC1 and OC2, during both cruises (Fig. 1).

During the spring cruise tidal conditions ranged between neaps (May 1) and springs (May 7, see Fig. 2b). During the fall cruise, on the other hand, the tides were of medium intensity, being weakest on October 25 and strongest on October 30 (Fig. 2c).

In this work, the AUV and winched profiler observations are used to validate the circulation model. In addition, data from the SATURN observation network endurance stations and National Oceanic and Atmospheric Administration (NOAA) tidal gauges are used to assess water elevations. All the observational data used in this study is available online (Sanford et al., 2015).

2.1.1. Endurance stations

Salinity and temperature measurements originate from five SATURN stations located in the lower estuary (Fig. 1): SATURN-

01 and the North Channel Bottom Node (NCBN1) in the North Channel, SATURN-03 in the South Channel and Jetty A (JETTA) near the mouth. In all these stations we use measurements of salinity and temperature near the bed, except for SATURN-03 where the observations are at three different depths 2.4, 8.6 and 13.0 m below datum. Tidal water levels are obtained from the NOAA tidal gauge located at Tongue Point (TPOIN).

2.1.2. Winched profiler

The winched profiler is a multi-instrument platform operated from a ship-mounted winch. The profiler was operated continuously, covering depths from near surface (roughly 2 m below free surface) to near bed (2 m above bed). Each vertical one-way profile took roughly 10 min.

The winched profiler carried a wide variety of instruments, only some of which will be discussed here. Salinity, temperature and pressure were measured with a SBE 37 CTD (Seabird-Electronics, Inc.) unit. The depth of the profiler was deduced from the CTD pressure data. Flow velocity relative to the profiler body was recorded with a Sontek ADV, and rotated to geographical coordinates using a magnetic compass (Spartan 3003D) and inertial sensor (BEI MotionPak II) data. All measurements were low-pass filtered to roughly 0.3 Hz frequency.

The winched profiler operations are shown as gray rectangles in Figs. 2(b) and (c), marked by either the OC1 or OC2 ship location. During the spring cruise the tidal cycle was sampled only partially: the profiler was not operated during the strongest ebb currents. Most of the data is from the OC2 ship location. In the fall cruise, when the currents were weaker, the profiler was operated continuously at both locations over several days. The tides were slightly weaker when the OC1 data set was collected. At these sites the difference in tidal amplitude due to the ship location is negligible compared to the temporal variability.

2.1.3. AUV fleet

To complement the winched profiler observations, two REMUS-100 autonomous underwater vehicles (Hydroid Inc.) were operated during the campaigns. The AUVs were equipped with SBE 49 CTDs (Seabird-Electronics, Inc.) for measuring water temperature and salinity and upward/downward looking ADCPs for currents. Water currents were calculated by transforming the raw ADCP data from vehicle coordinates to Earth coordinates using the vehicle's tilt and compass information.

The AUVs navigate underwater using acoustic transponder pairs. While underwater, the real-time position of the vehicle is estimated from multiple data sources, including the acoustic range to nearest transponders, speed over ground, and relative currents. The AUV position is eventually corrected in post-processing to determine the trackline that best matches with GPS fixes obtained during the mission. The temporal resolution of the final ADCP data

was roughly 1 Hz. The renavigated CTD data had a sampling rate of roughly 9 Hz. For the purposes of this paper, the CTD data was binned to 0.5 Hz temporal resolution.

A mission planner, informed by the SATURN daily forecasts of baroclinic circulation, was used to design the timing and routes of the AUVs in the strong currents. The AUVs were programmed to sample both along and cross-channel sections in the North Channel, covering depths from roughly 2 m below surface to 2 m above the bed. Both along and cross-channel missions were repeated on different days. The AUVs were only operated during floods to avoid the strongest currents.

The time span of each mission is illustrated in Fig. 2(b) and (c) with a gray vertical bar and the corresponding mission number. In this paper the along-channel missions M44 (high flow, neap tides), M48 (high flow, spring tides) and M70 (low flow, medium tides) are examined in detail. All the missions, however, contribute to the error metrics show in Tables 1, 2 and A.3.

2.2. Simulations

2.2.1. Circulation model

Numerical simulations were carried out with the unstructured grid, finite element model SELFE (Zhang and Baptista, 2008). SELFE uses triangular elements in the horizontal that are extruded in vertical to form a 3D prismatic mesh. Linear continuous and non-confirming basis functions are used for discretizing free surface elevation and horizontal velocity, respectively. Tracers are constants within each element. The vertical grid consists of terrain following and free surface adapted S grid (Song and Haidvogel, 1994) near the surface, and equipotential z grid below. In this work the transition depth from S- to z-levels is 100 m. The thickness of the z levels range from 2300 m in the deepest part of the domain to 5 m near the transition to S coordinates. The equations are solved in Cartesian x, y, z space instead of true vertical coordinate space.

SELFE implements an implicit free surface equation with wetting and drying. Vertical diffusion is treated implicitly, and advection of momentum is marched in time with an Eulerian–Lagrangian (ELM) method. The tracer transport scheme is based on a mass conservative, first order, finite volume upwind method with slope limiters. This formulation circumvents the most restrictive constraints on Courant number allowing relatively long time steps without affecting numerical stability. The only exception is the tracer transport scheme where time steps are constrained by the CFL stability condition of the upwind scheme, to retain conservation and monotonicity properties. This time step is chosen at run time, based on the local velocity and tracer fields, and element size. A stable time step is determined for each element, and the global minimum value is used in the transport algorithm.

Vertical subgrid-scale mixing is described by the Generic Ocean Turbulence Model (GOTM, Burchard et al., 1999), which provides

Table 1
Root mean square error of simulated salinity for different grid and time step configurations. The Murphy (MS) and Willmott skill (WS) scores are given in parenthesis. The coarse grid performs better with longer time steps, while the opposite is true for the finest grid: best overall performance is obtained with 36 s time step. Station error is computed with observations during the spring (May 1–May 19) and fall periods (October 17–November 3) for the following stations: SATURN-03 (depths 2.4, 8.6 and 13.0 m below datum), SATURN-01 (19.5 m), Jetta (6.4 m), NCBN1 (12.0 m). Station errors are presented also excluding SATURN-01 station. AUV and winched profiler error includes all the collected data. Minimum RMSE is indicated with bold face. Willmott score is in italics.

Setup		RMSE [psu] (MS,*WS*)			
Grid	Δt (s)	Stations		Winched Prof.	AUV
		All	Excl. SATURN-01		
Coarse	90	10.46 (0.02, *0.76*)	7.00 (0.54, *0.88*)	7.09 (0.52, *0.89*)	5.09 (0.79, *0.94*)
Coarse	60	10.59 (−0.00, *0.76*)	6.93 (0.55, *0.88*)	7.35 (0.48, *0.88*)	5.13 (0.79, *0.94*)
Coarse	36	13.85 (−0.71, *0.64*)	10.62 (−0.06, *0.75*)	13.53 (−0.76, *0.66*)	10.45 (0.11, *0.74*)
Fine	90	11.12 (−0.10, *0.73*)	8.39 (0.34, *0.83*)	7.77 (0.42, *0.86*)	6.63 (0.64, *0.89*)
Fine	60	9.25 (0.24, *0.80*)	6.40 (0.62, *0.90*)	6.01 (0.65, *0.92*)	4.59 (0.83, *0.95*)
Fine	36	**7.57** (0.45, *0.86*)	**4.84** (0.77, *0.94*)	**5.05** (0.76, *0.94*)	**3.68** (0.89, *0.97*)

Table 2
Skill metrics for the best model configuration. The metrics of water levels, salinity and temperature are evaluated at the given station over the two cruise periods. *AUV-WP* denotes average skill on all AUV and winched profiler data. Bias stands for mean error and standard deviation; MS is the Murphy score; WS is the Willmott score. Station depths are as in Table 1 except for SATURN-03 where the deepest value is used (13 m).

Variable	Data set	Bias	RMSE	MS	WS
Water level	TPOIN	-0.18 ± 0.14 m	0.23 m	0.92	0.98
Salinity	JETTA	1.33 ± 4.37 psu	4.57 psu	0.81	0.95
Salinity	SATURN-01	-8.25 ± 9.87 psu	12.86 psu	-0.71	0.68
Salinity	SATURN-03	-2.51 ± 3.93 psu	4.66 psu	0.78	0.94
Salinity	AUV-WP	-0.53 ± 4.76 psu	4.79 psu	0.81	0.95
Temperature	JETTA	-0.34 ± 0.56 °C	0.65 °C	0.58	0.89
Temperature	SATURN-01	0.38 ± 1.19 °C	1.25 °C	-0.10	0.67
Temperature	SATURN-03	-0.14 ± 0.76 °C	0.78 °C	0.46	0.82
Temperature	AUV-WP	-0.28 ± 0.37 °C	0.47 °C	0.70	0.93

Table A.3
Root mean square error of simulated salinity for different values of bottom roughness length. All simulations were run with the fine grid and 36 s time step. The Murphy (MS) and Willmott skill (WS) scores are given in parenthesis. The station, AUV and winched profiler data sets are defined as in Table 1. Minimum RMSE is indicated with bold face; Willmott scores are in italics.

z_0 [m]	RMSE [psu] (MS, *WS*)			
	Stations		Winched Prof.	AUV
	All	Excl. SATURN-01		
1.0×10^{-3}	7.79 (0.42, *0.86*)	5.02 (0.75, *0.94*)	5.12 (0.75, *0.94*)	3.60 (0.89, *0.97*)
5.0×10^{-4}	7.72 (0.43, *0.86*)	4.97 (0.76, *0.94*)	5.08 (0.75, *0.94*)	**3.56** (0.90, *0.97*)
1.0×10^{-4}	7.57 (0.45, *0.86*)	**4.84** (0.77, *0.94*)	**5.05** (0.76, *0.94*)	3.68 (0.89, *0.97*)
1.0×10^{-5}	**7.56** (0.45, *0.86*)	4.91 (0.76, *0.94*)	5.34 (0.73, *0.94*)	4.49 (0.84, *0.96*)

vertical eddy viscosity and diffusivity to the circulation model. In this work a standard k-ε turbulence closure model with Canuto A stability functions (Canuto et al., 2001) is used. The minimum value for eddy viscosity and diffusivity was set to 1×10^{-9} m^2 s^{-1}. No explicit horizontal diffusion is applied.

All simulations were carried out on Texas Advanced Computing Center (TACC) Stampede cluster, a part of the Extreme Science and Engineering Discovery Environment (XSEDE, Towns et al., 2014).

2.2.2. Setup and forcings

The model domain covers a part of the Northeast Pacific continental shelf, extending from latitude 39N to 50N and roughly 300 km in the offshore direction (Fig. 3). The vertical S grid consists of 37 levels and is defined such that in shallow areas (below 30 m, i.e. most of the estuary) the coordinates revert to conventional sigma layers. In the deepest part of the continental shelf there are 17 z-levels resulting in a total 54 vertical levels.

Bathymetry is a composite of multiple National Geophysical Data Center data sets: ETOPO2v2 (NGDC, 2006), 3 arc second Coastal Relief Model (NGDC, 2011), as well as 1/3 arc second rasters for selected coastal regions. Oregon Department of Geology and Mineral Industries data is used in the estuary. Bathymetry in the lower estuary and navigation channel is corrected using recent US Army Corps of Engineers survey data.

Temperature, salinity and water elevations are imposed at the Pacific boundary from global Navy Coastal Ocean Model (NCOM, Barron et al., 2006) simulations. NCOM provides only subtidal water elevation, on top of which 8 dominant tidal constituents (O1, K1, Q1, P1, K2, N2, M2, S2) are superimposed. The tidal constituents are obtained from a regional inverse model (Myers and Baptista, 2001). In the proximity of the open boundary (50 km), temperature and salinity values are nudged towards NCOM values using a relaxation time of 2 days. In the riverine end of the grid, water flux and temperature are imposed at Beaver Army, 86 km upstream of the mouth, using US Geological Survey (USGS) gauge data. Salinity is set to zero.

Atmospheric forcing originates from the NOAA/NCEP North American Mesoscale Forecast System. Forecast resolution is 12 km. Wind speed 10 m above surface, air pressure, and heat radiation fluxes are used. Evaporation and precipitation were neglected as they are insignificant compared to the riverine freshwater flux.

Two time periods corresponding to the cruises, ranging from 1 May through 20 May 2012 (spring) and 25 October through 3 November 2012 (fall), were simulated. In both cases the model was spun up for 7 days before the analysis period. Considering the low residence time of the estuary (typically between 1 and 4 days), a 7 day warm-up period is sufficient for establishing circulation in the estuary. Initial conditions for salinity and temperature were obtained from NCOM in the shelf, while constant values were used in the estuary (0 psu for salinity and 10 °C for temperature).

Bottom stress is parametrized by a conventional law-of-the-wall condition, where the bottom roughness length z_0 was tuned to match salinity observations. For the sake of simplicity z_0 was kept constant in space and time. All the presented runs use value $z_0 = 1 \times 10^{-4}$ m, unless otherwise specified.

2.3. Skill metrics

The basic metric of model skill used in this paper is the root mean square error (*RMSE*),

$$RMSE = \sqrt{\langle (m - o)^2 \rangle},$$

where $o = \{o_i\}_{i=1}^n$ and $m = \{m_i\}_{i=1}^n$ are the observed and modeled time series, respectively, and $\langle \cdot \rangle$ denotes the average over the series.

RMSE is an intuitive metric as it has the same unit as the primary variable, but it does not allow comparing the skill of different variables. To this end we also use the nondimensional Willmott score (*WS*, Willmott, 1981), defined by

$$WS = 1 - \frac{\langle (m - o)^2 \rangle}{\langle (|m - \langle o \rangle| + |o - \langle o \rangle|)^2 \rangle}.$$
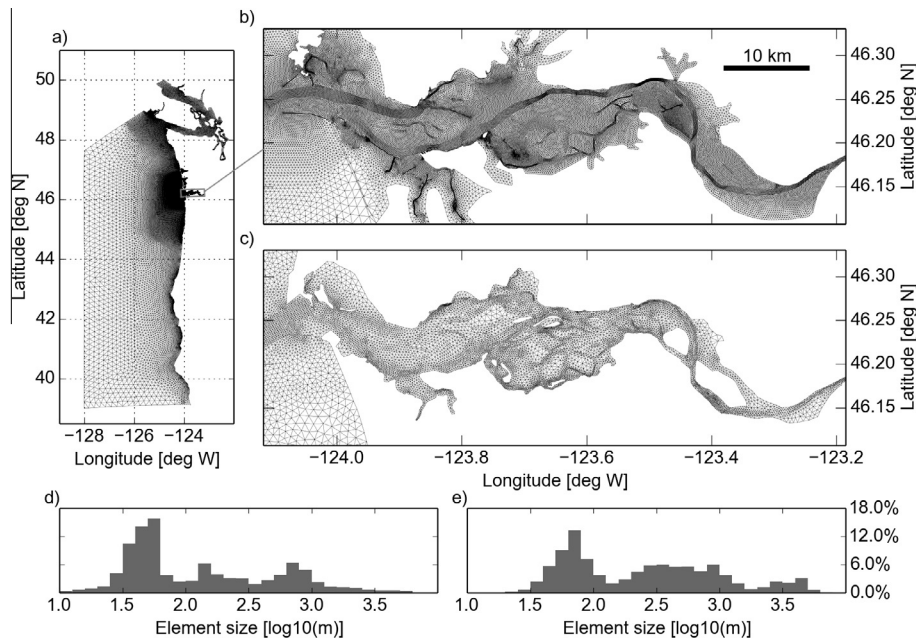
Fig. 3. (a) Horizontal mesh for the full domain (fine grid), consisting of 109 000 triangles and 56 000 nodes; (b) close up view of the estuary (fine grid); (c) close up view of the estuary (coarse grid). The coarse grid covers the same full domain as the fine grid and consists of 39 000 triangles and 21 000 nodes.The three-dimensional grids have roughly 2.9 and 1.0 million prisms, respectively. The histograms in (d) and (e) show the triangle inradius in meters on logarithmic scale for the fine and coarse grid, respectively.

The Willmott score is used in several modeling studies, including Liu et al. (2009) and Elias et al. (2012) for the Columbia River estuary. Liu et al. (2009) computed WS separately for tidal and synoptic time scales of time series. In the case of CTD casts, they computed WS for each depth and then averaged the results over either the surface or bottom layer of the shelf in order to obtain more reliable skill estimates.

WS has been criticized for being relatively insensitive to model parameters and producing high skill values for entirely uncorrelated signals (Ralston et al., 2010). Therefore we also use the Murphy score (MS, Murphy, 1988)

$$MS = 1 - \frac{\langle (m-o)^2 \rangle}{\langle (m_r - o)^2 \rangle},$$

where m_r is the output of a reference model to compare against. The Murphy score is 1 for a perfect model, zero for a model that is equivalent to the reference model, and negative for a model worse than the reference.

Here the reference model is taken as the mean of the observations, $m_r = \langle o \rangle$. With this definition, the metric can be written as $MS = 1 - NMSE$, where NMSE is the normalized mean square error, i.e., mean square error divided by the variance of the observations ($Var(o)$). Similarly, MS can be expressed in terms of $RMSE : MS = 1 - RMSE^2/Var(o)$. In this paper MS is the primary adimensional error metric, while WS is only presented to allow comparison to previous studies.

3. Model-data comparison

3.1. Sensitivity to mesh resolution and time step

The skill of circulation models strongly depends on grid resolution and time step, the latter being restricted either by stability or accuracy criteria. A brief sensitivity study was carried out using two different grids: a coarse grid, consisting of 39 000 triangular elements, and a fine grid with 109 000 elements (Fig. 3b and c). The fine mesh was designed to better represent the dominant

topographic features of the estuary while keeping computational cost manageable. Vertical discretization was the same in both cases. Horizontal resolution in the main channels is roughly 300 m and 180 m for the two grids, respectively. These grids were tested with three different time steps: 90, 60 and 36 s. This time step applies to the free surface equation and internal 3D mode. The time step for tracer transport, chosen at run time, was 5 s and 35 s on average for the fine and coarse grids, respectively.

Fig. 4 presents a comparison of bottom salinity time series at selected stations in the North and South Channel. Root mean square errors as well as Willmott and Murphy skill scores of modeled salinity are listed in Table 1 with separate scores for station, AUV and winched profiler data sets. Note that the Willmott score is fairly high in all cases, which could be misleading. The Murphy score, on the other hand, captures a broad range of values, including negative ones indicating poor skill.

The coarse grid performs moderately well with 90 s time step. The semi-diurnal salinity intrusion is well captured (Fig. 4), but salinity is underestimated during neaps and weaker floods especially at North Channel stations NCBN1 and SATURN-01. Decreasing the time step does not improve performance, in fact the RMSE errors grow rapidly as time step decreases. The fine grid with 90 s time step is slightly worse than the coarse grid. With shorter time steps, however, the skill improves significantly: the combination of fine grid and 36 s time step clearly yields superior performance. Salt intrusion at weaker floods is much better represented than with the coarse grid. Decreasing the time step further did not result in notable improvement (not shown).

The results indicate that for a given grid the optimal skill is obtained at a specific temporal resolution. ELM methods are known to have an optimal time step, typically corresponding to mesh Courant number $Cu > 1$ (Baptista, 1987). The mesh Courant number for horizontal advection is given by $Cu = (U\Delta t)/\Delta x$, where U is the magnitude of characteristic velocity, Δt is the time step and Δx is the horizontal element size. Reducing time step so that $Cu < 1$ typically increases numerical diffusion in ELM schemes, deteriorating performance. In the presented simulations, the Courant number for horizontal advection is always greater than one in
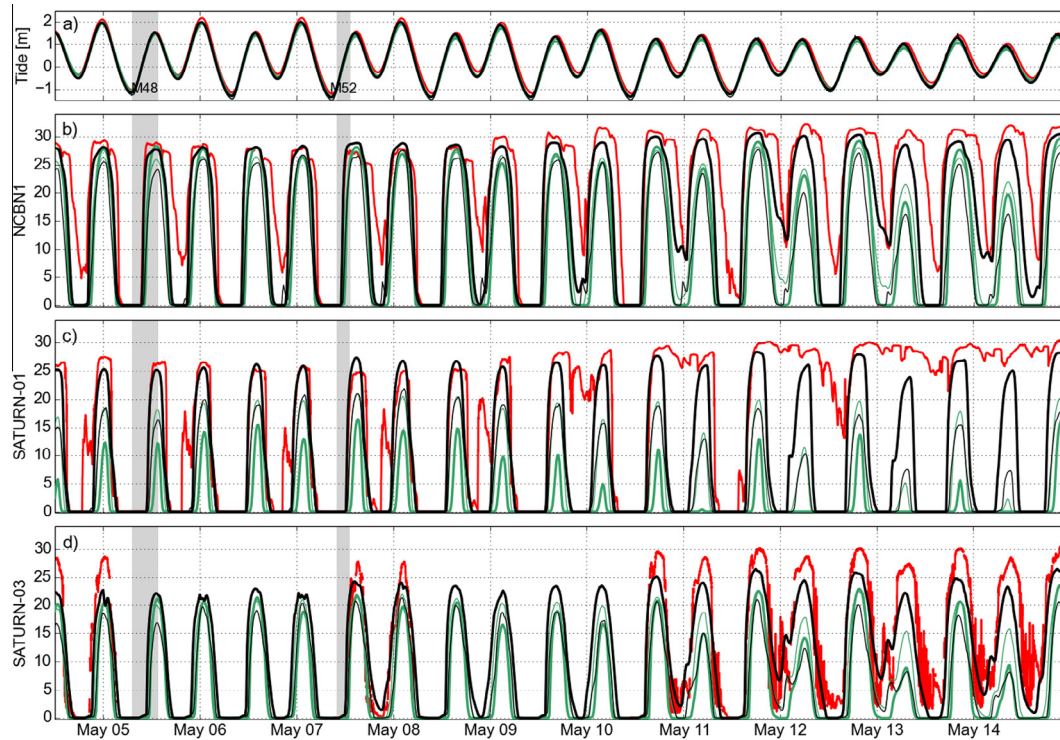
Fig. 4. Comparison of bottom salinity at endurance stations for the spring cruise period. Each panel compares model outputs to the observations (red line) for three stations, NCBN1 (at depth 12.0 m, panel b), SATURN-01 (13.0 m, panel c) and SATURN-03 (19.5 m, panel d). Four model setups are plotted: fine grid (black line), coarse grid (green line), 36 s and 90 s time step (thick and thin lines, respectively). Panel (a) shows comparison of water elevations at TPOIN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the lower estuary. For comparison, the time step corresponding to $Cu = 1$, as determined by the smallest element in the domain, would be roughly 5 s for the fine grid. The ELM method therefore reduces the computational cost by allowing longer time steps.

Due to the large differences in performance, we are only considering the best model configuration, fine grid with 36 s time step, for the remainder of the paper. Even in this case, however, the model does not capture the intertidal salt retention at SATURN-01 during neap tides (Fig. 4c). This is a known limitation of the model, and highlights special characteristics of the SATURN-01 station: the bathymetry around the station is deeper than the North Channel in general (Fig. 1), which can prevent salt from being flushed downstream when ebb currents are weak near the bed. The model has difficulty in capturing these local dynamics. Because the mesh is fine enough to resolve the bathymetry in this region, this shortcoming is likely related to underestimated gravitational circulation in the model, which weakens the landward salt intrusion in the bottom layer.

The ribbon diagram in Fig. 5 illustrates low pass filtered error at SATURN-01 for the entire year 2012, obtained from a baseline hindcast simulation. The hindcast simulation is comparable to the fine-grid model presented here, although the domain extends upstream to Bonneville dam and also covers a part of the Willamette River. The error is plotted against the two main forcings of the system: tidal range, estimated from TPOIN water elevations, and low-pass filtered river discharge at Beaver Army terminal. The poorest skill coincides with high flow and neap tide conditions, i.e., when the estuary is a salt-wedge system and stratification is strongest. Skill is highest for spring tides regardless of the river flow regime.

3.2. Sensitivity to bottom friction

Error metrics for the best model (fine grid, 36 s time step) with varying bottom roughness length (ranging from 1×10^{-5} to
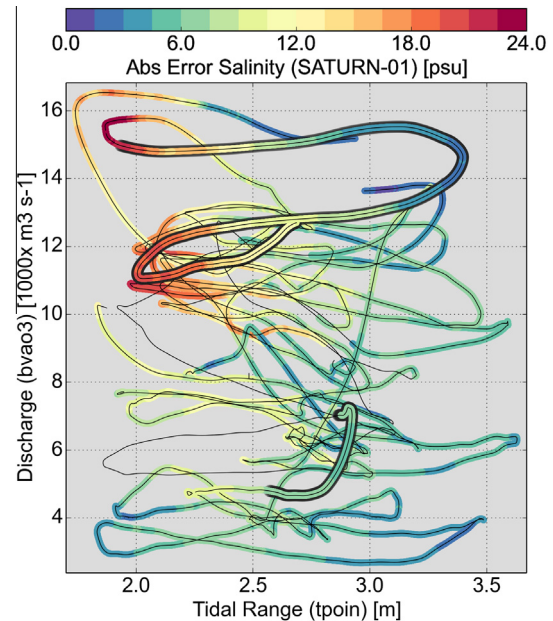
Fig. 5. Error in modeled salinity at SATURN-01 for the entire year 2012. Low pass filtered absolute error is plotted against low pass filtered river discharge (at Beaver Army terminal) and tidal range (obtained from TPOIN tide gauge). The results are obtained from a year-long baseline simulation, which uses the same grid resolution and time step as the best model presented herein. The error is largest under high flow and neap tide conditions when the estuary is a salt wedge system. At spring tides the model performs clearly better. The two cruise periods are indicated with dark shading (top, spring cruise; bottom, fall cruise). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1×10^{-3} m) are shown in the Appendix (Table A.3). The sensitivity of the skill to bottom friction is much smaller than sensitivity to

the time step, to the extent that the Willmott score does not differentiate the skill of the tested bottom roughness length values.

Based on the skill metrics, we have chosen the value $z_0 = 1 \times 10^{-4}$ m, also supported by visual comparison against AUV and winched profiler data sets. For comparison, Elias et al. (2012) used 2D Chézy bed roughness coefficients between 50 and 61 $m^{1/2}$ s^{-1}. Assuming 10 m water column depth and 20 σ-layers, those values correspond to roughness lengths 8×10^{-3} and 2×10^{-3} m, respectively. MacCready et al. (2009) used a drag coefficient 3×10^{-3}, which, under the same assumptions, corresponds to roughness length 3×10^{-4} m, similar to the value used here.

3.3. Overall skill metrics

Due to the large amount of data it is not practical to present a direct model-observation comparison of each individual data set. In this section some integrative error metrics are presented to demonstrate overall model performance during the two cruise periods. In the subsequent sections some illustrative individual cases are shown in more detail.

Table 2 presents error metrics for water levels, temperature and salinity at various stations. In general, water levels are simulated with better Murphy score (0.92) than salinity (0.81 at best station) or temperature (0.70). The Willmott scores are again high in all cases.

Skill is clearly worst at SATURN-01, which highlights the distinct nature of this station. Considering the combined AUV and winched profiler data set, mean error is −0.5 ± 4.8 psu for salinity and −0.3 ± 0.4 °C for temperature, showing that the model is not strongly biased in the North Channel (except at SATURN-01). Some caution is needed in interpreting these bulk statistics, however, because the AUV data covers predominantly floods rather than ebbs.

The presented model skill is similar to results found in Liu et al. (2009) and Elias et al. (2012). Neither of these studies included a station comparable to SATURN-01, which is consequently omitted from the comparison. Liu et al. (2009) simulated the estuary-plume-shelf system for the summer 2004, and reported Willmott skill score separately for low and high pass filtered components. Water levels at Astoria had skill 0.70 and 0.97 for the two components, respectively. Salinity and temperature skill was assessed at three different stations in the estuary. The scores were 0.72 and 0.90 for salinity, and 0.71 and 0.90 for temperature. All of these measures are slightly lower than the Willmott scores presented here. It is worth noting, however, that the focus of Liu et al. (2009) was on the plume rather than the estuary. Elias et al. (2012) focused on the circulation near the mouth of the Columbia for fall 2005. They reported a better Willmott score 0.99 for elevations at Astoria. Salinity skill in the South Channel (station red26, close to SATURN-03) was 0.90, again slightly less than in the present study. It should be noted, however, that the skill metrics are not directly comparable as the simulated time periods were different in these studies.

Histograms of observed salinity from the AUV and winched profiler are presented in Fig. 6, with comparison to modeled values. For the spring (high flow) conditions, the observed salinity distribution is highly polarized: There is a large fraction (>30%) of fresh water and a peak of saline water at about 27 psu. Intermediate salinities (between 10 and 21 psu) are less abundant, indicating strong separation between the two end member water masses. In the fall (low flow) conditions, on the other hand, partially mixed water masses dominate with peaks around 7 and 28 psu. Fresh water fraction is very small due to smaller river flow and stronger mixing. There is a larger fraction of highly saline water implying stronger
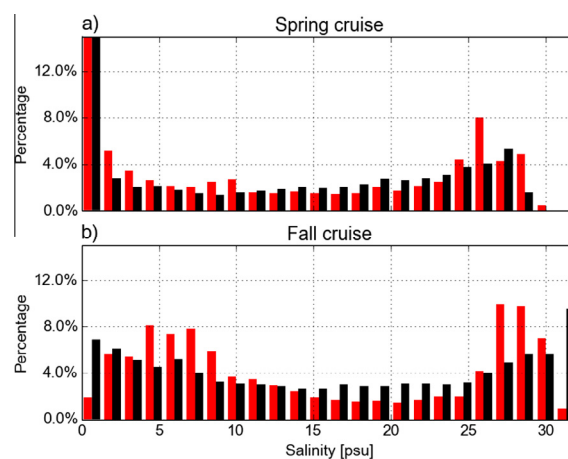
Fig. 6. Histogram of observed and simulated salinity for (a) spring and (b) fall cruise periods. The histogram contains all observation data points from the AUV and winched profiler missions (red), and corresponding data set extracted from the model outputs (black). The fresh water fraction in (a) is greater than the shown range, exceeding 30% for both the observations and model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

salinity intrusion. The two water masses are clearly separated in this case as well.

The model produces similar bimodal distribution in both cases, but shows more intermediate salinities indicating stronger mixing. Furthermore, under low flow conditions the model tends to overestimate maximum salinity.

3.4. AUV observations

AUV observations provide a view into the spatial distribution of salinity in the North Channel. Along-channel transects obtained under high flow and neap tide conditions (AUV mission M44) are compared in this section to spring tides (M48) and low flow conditions (M70).

Mission M44 is the only extensive AUV data set collected under high flow and neap tide conditions. Selected salinity transects are presented in Fig. 7. In the first transect (panel b) the salt wedge arrives around 7:00 PST, roughly in the middle of the rising tide (see panel a). Simulated salinity field, interpolated to the same points in space and time, is shown in panel (c). The model correctly reproduces the incoming salt wedge, but it is delayed by roughly 1 km or 15 min. A similar delay is visible in the subsequent transect (panel f). Note the extremely sharp halocline in the observations, particularly in panel (h).

The model captures the main characteristics of the salt wedge, with two notable caveats: the halocline is smoothed out and salinity in the bottom layer is underestimated. Maximum observed salinity near the bed is around 31 psu while the modeled values do not exceed 25 psu.

Fig. 8 shows along channel currents for the same transects of mission M44. The inflowing bottom current and slower surface layer are visible in the observations (panels b and e). In the last transect (h) the tides are turning, the surface layer already starting to ebb. Maximum flood velocity occurs right beneath the halocline. The subsurface velocity maximum, collocated with the halocline, is characteristic of strongly stratified estuaries (Geyer and Farmer, 1989) and implies that bottom boundary layer growth is limited by stratification (Stacey and Ralston, 2005; Ralston et al., 2010).

The model reproduces similar flow characteristics: flood currents are stronger in the bottom layer and attain maximum near the halocline. The velocity field, however, is smoother than observed.
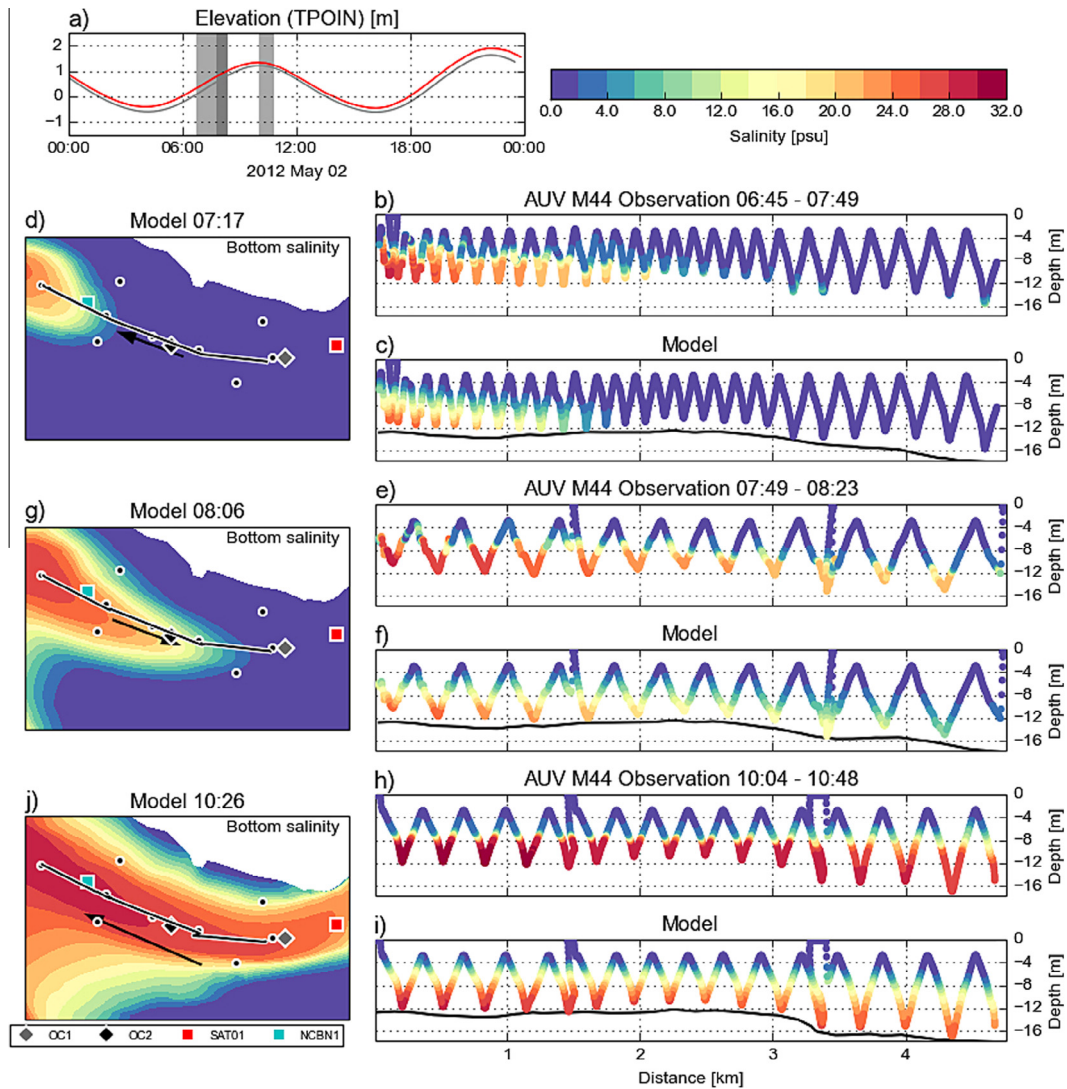
Fig. 7. Recorded salinity from AUV mission M44 versus the model output. Panels on the right (b, e, h) show the observed salinity and corresponding model output (c, f, i) for different AUV transects. Panels on the left (d, g, j) show the AUV track superimposed on simulated bottom salinity field. The black arrow indicates the AUV travel direction. Panel (a) presents the tidal elevation at Tongue point for the shown time period (red, observed; gray, simulated). Vertical bars indicate the time span of the shown AUV transects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The mission M48 is used as an example of high flow and spring tide conditions. Salinity fields are presented in Fig. 9. Note that in this case the AUV operated only in the downstream portion of the transect. Compared to neap tides, the observed halocline is thicker due to stronger tidal mixing. The model captures the timing of the salt wedge arrival into the North Channel better than under neap tide conditions. Maximum salinity is underestimated in this case as well (panel f), but the difference is smaller: 27 psu versus observed 31 psu. Simulated salinity field is more diffuse than the observations.

Mission M70 is an example of low flow and medium tide conditions (Fig. 10). The first transect corresponds to an earlier phase of the tide than in Figs. 7 and 9, but saline waters are already present in the bottom layer throughout the transect. This is consistent with the earlier notion that low river discharge, about 4000 m³ s⁻¹, is not enough to flush all the salt from the North Channel. There are clearly more intermediate salinity classes, consistent with the histogram (Fig. 6b). In contrast to the high flow conditions the AUV observations show a continuous salinity gradient across the water column. As the flood progresses, however, stratification increases and the saline bottom layer becomes nearly homogeneous.

The model performs better under the low flow conditions: because the density gradients are smoother, the high diffusivity of the model is less of an issue than under high flow conditions. The observed sharp halocline, however, is not captured.

3.4.1. Estimating bulk diffusivity of the model

In order to estimate the bulk diffusivity of the numerical model, we fitted an analytical solution of a 1D vertical diffusion equation to the AUV salinity data. To avoid interference from physical mixing processes, such as internal waves and shear instabilities, data from AUV mission 44, that exhibits the sharpest observed halocline, was used. The details of the fit are presented in Appendix B.

Assuming that the salinity field was initially a sharp step function that has been diffused over a limited time, one can estimate the bulk diffusivity for both the observations and model. Following the analysis, the bulk diffusivity of the observations was 4×10^{-5} m² s⁻¹ while in the model it was a magnitude higher, 5×10^{-4} m² s⁻¹. During this time period, turbulent eddy diffusivity in the model (produced by the turbulence closure) was 1×10^{-9} m² s⁻¹ in the North Channel where stratification was
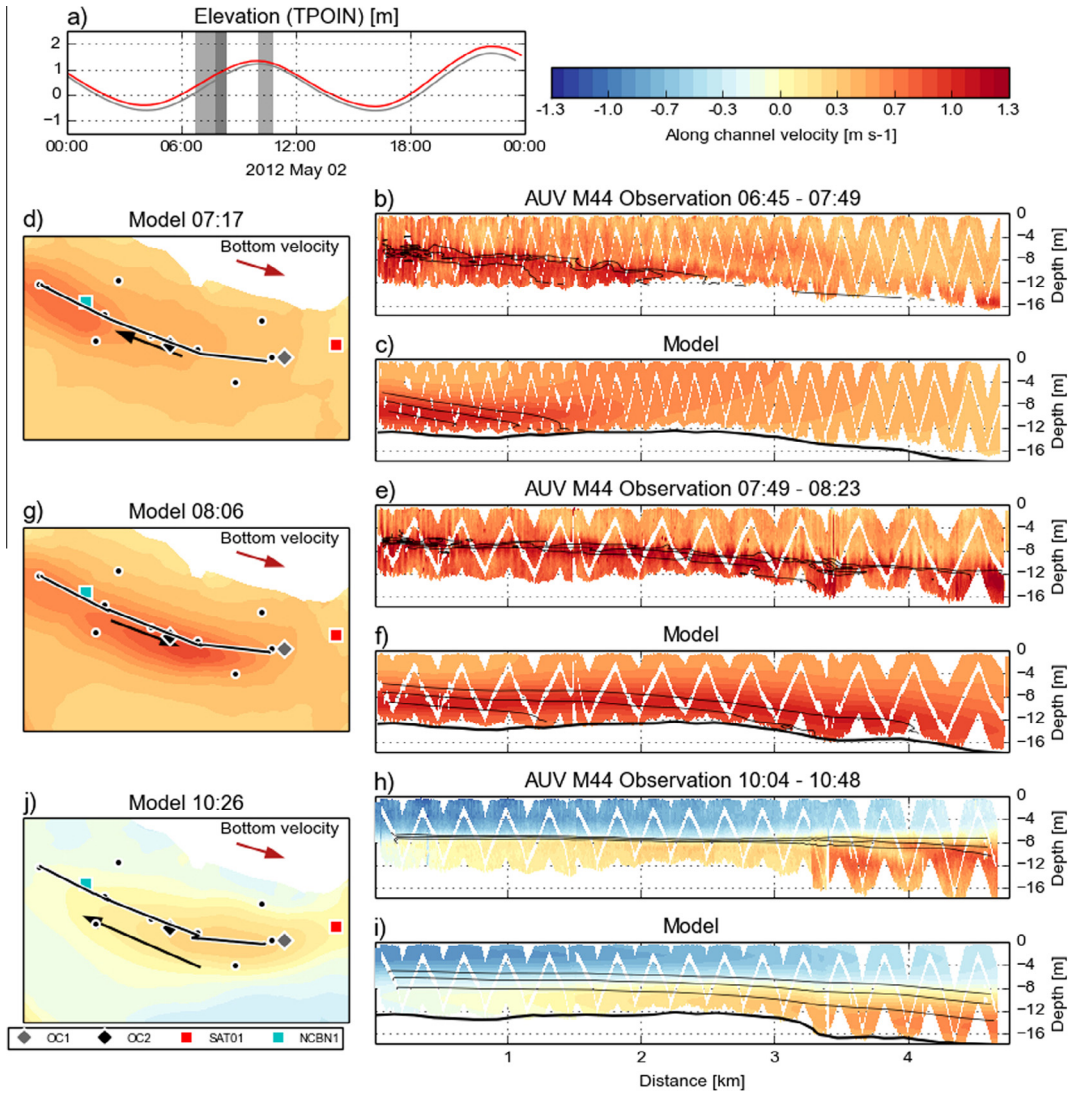
Fig. 8. Recorded along-channel velocity from AUV mission M44 versus the model output. Panels on the right (b, e, h) show the observed velocity and corresponding model output (c, f, i) for different AUV transects. Thin superimposed lines indicate 10, 16 and 22 psu isohalines. Panels on the left (d, g, j) show the AUV track superimposed on simulated bottom velocity field (evaluated one grid level above the bed). The black arrow indicates the AUV travel direction. Along channel velocity is defined along the red arrow. Panel (a) presents the tidal elevation at Tongue point for the shown time period (red, observed; gray, simulated). Vertical bars indicate the time span of the shown AUV transects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

present. It was higher only near the bed where the bottom layer is nearly homogeneous, reaching up to 1×10^{-2} m^2 s^{-1}. This indicates that mixing is indeed considerably larger in the model, and the destruction of stratification is not related to the parametrization of turbulence, but rather inherent to the numerics of the model.

3.5. Winched profiler observations

The winched profiler data set, recorded over multiple tidal cycles, provides a more comprehensive view on the tidal dynamics. During the spring cruise most of the observations were collected at the downstream ship location OC2. The profiler was not operated at major ebbs and only covered mid-depths of the water column. Fig. 11 shows the observed salinity at OC2 for three consecutive tidal cycles, corresponding to high flow and spring tide conditions. The tidal range was increasing throughout the campaign, reaching maximum on May 7.

Although the coverage is sparse, the observations suggest that the water column is mostly fresh, and significant salinity is only

observed in the bottom layer in the end of flood and early ebb tides. This is consistent with the AUV observations (Fig. 7) and the station time series (Fig. 4). In Fig. 11 the model reproduces the timing and shape of the salinity intrusion relatively well. As in the case of AUV comparisons, the most salient difference is the over-diffused salinity gradients. In the first tidal cycle salinity intrusion is underestimated, while slight overestimation is visible in the last flood phase on May 7 when the tides were the strongest.

The winched profiler observations on May 5 show that some salt is retained near the bottom during lesser ebbs. In Fig. 4 similar retention is seen in SATURN-01 for the same time period (May 5 to 7). The model does not capture the salt retention in either case, suggesting that the simulated baroclinic landward bottom current is weaker than in reality. This is likely a consequence of numerical mixing as well: diffusion reduces not only the stratification but also the along-channel density gradient due to the tilt of the halocline. Therefore the seaward barotropic density gradient exceeds the landward baroclinic density gradient during ebbs, and pushes the entire water column seaward.

During the fall cruise the profiler was operated continuously over multiple days. OC1 observations were obtained at weaker
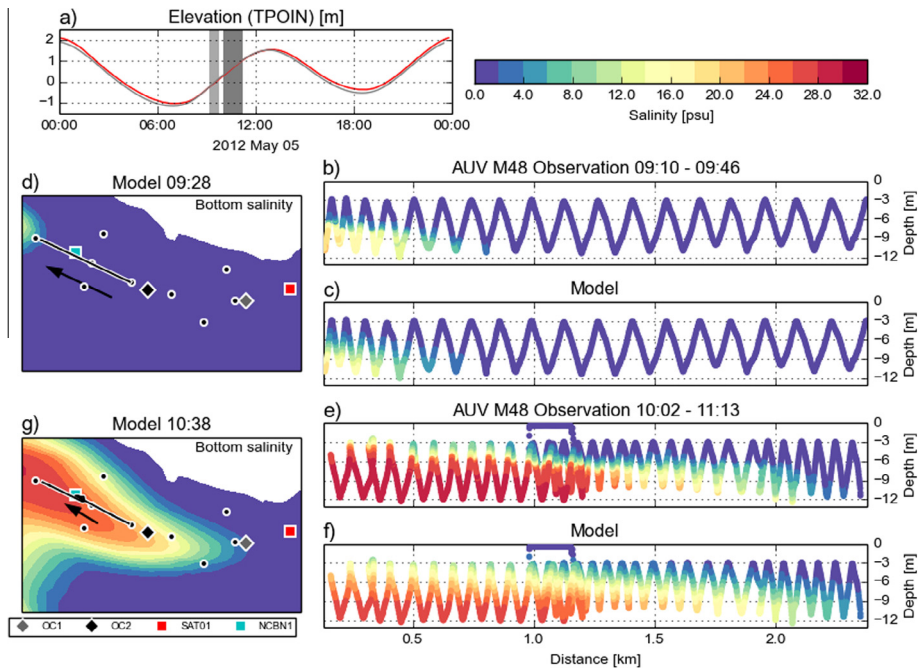
Fig. 9. Recorded salinity from AUV mission M48 versus the model output. Panels on the right (b, e) show the observed salinity and corresponding model output (c, f) for different AUV transects. Panels on the left (d, g) show the AUV track superimposed on simulated bottom salinity field. The black arrow indicates the AUV travel direction. Panel (a) presents the tidal elevation at Tongue point for the shown time period (red, observed; gray, simulated). Vertical bars indicate the time span of the shown AUV transects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
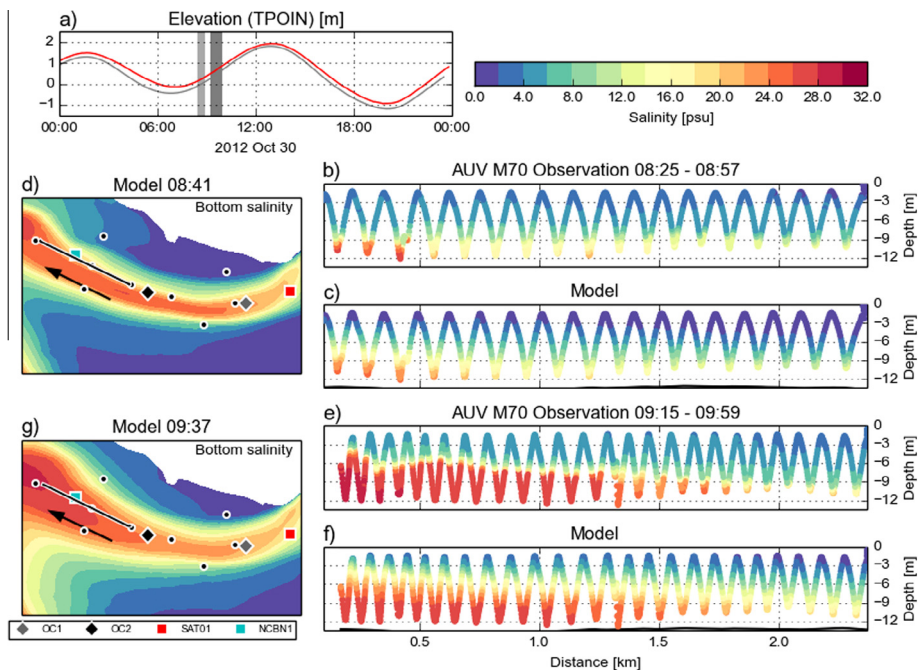
Fig. 10. Recorded salinity from AUV mission M70 versus the model output. Panels on the right (b, e) show the observed salinity and corresponding model output (c, f) for different AUV transects. Panels on the left (d, g) show the AUV track superimposed on simulated bottom salinity field. The black arrow indicates the AUV travel direction. Panel (a) presents the tidal elevation at Tongue point for the shown time period (red, observed; gray, simulated). Vertical bars indicate the time span of the shown AUV transects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)xmllabelf0050

tides than the OC2 data set (Fig. 2). The difference in tidal range is only 0.4 m, but it has impact on the salt dynamics in the North Channel.

Figs. 12 and 13 present salinity and along channel velocity for the OC1 and OC2 ship locations, respectively. Salt intrusion is stronger than under high flow conditions, the observed salinity regularly reaching 32 psu (Figs. 12b and 13b). The water column

is stratified over the entire tidal day and brackish waters are found at the bottom at all times, as suggested by the AUV mission M70 as well. Salt retention near the bed is strong in both data sets, bottom salinity often remaining above 28 psu.

Stratification and salt retention are, however, modulated by the tides. During the first day at OC1 (Fig. 12b), retention occurs during both the major and lesser ebbs. As the tides grow stronger in the
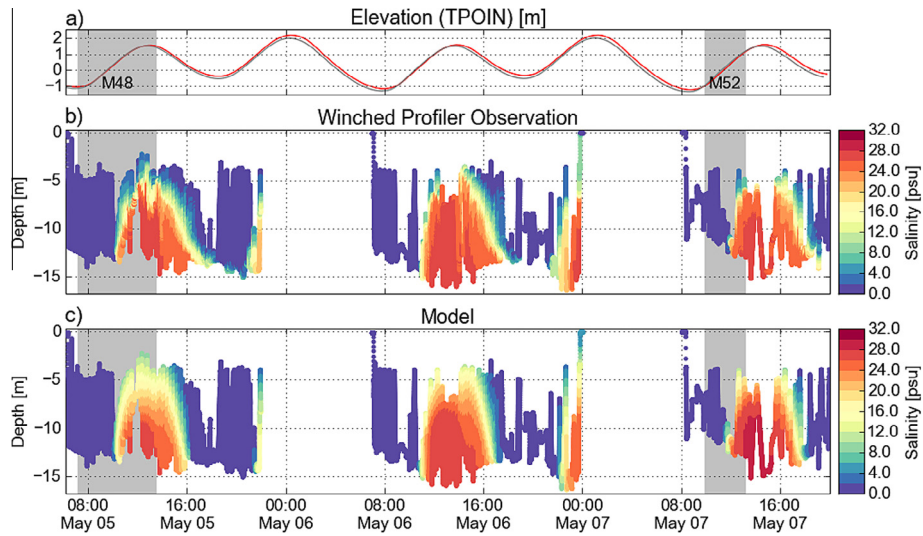
Fig. 11. Winched profiler recorded salinity at OC2 (panel b) versus model output (c), for high flow and spring tide conditions. Panel (a) shows water elevation at TPOIN station (red, observed; gray, simulated). The shaded areas indicate coinciding AUV missions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
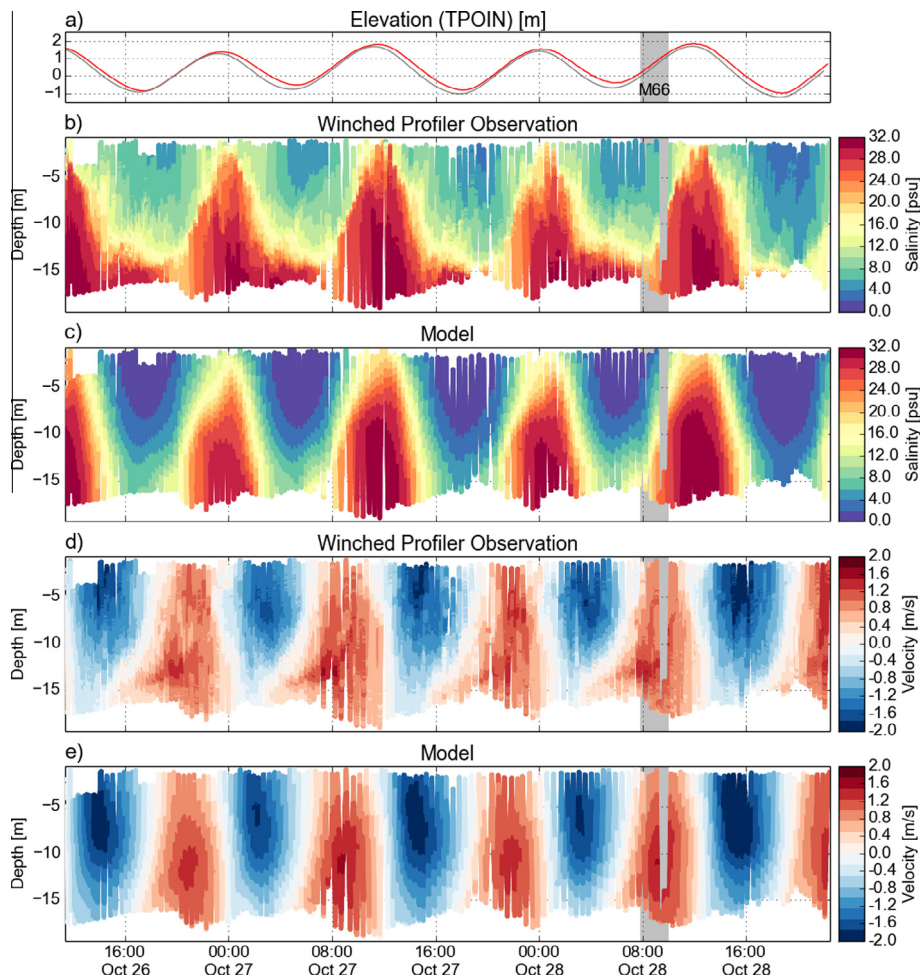
Fig. 12. Winched profiler observations at OC1 versus model output for low flow conditions. (a) water elevation at TPOIN station (red, observed; gray, simulated); (b) observed salinity; (c) modeled salinity; (d) observed along channel velocity; (e) modeled along channel velocity. Along channel velocity in (d) and (e) is taken as the principal component of each data set. The shaded area indicates coinciding AUV mission. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

subsequent tidal cycles, barotropic pressure gradient becomes stronger and salt retention decreases during major ebbs. At the same time the surface layer becomes fresher. At OC2 (Fig. 13b) a similar tendency is visible; high near-bed salinity values are observed during lesser ebbs, but not during major ebbs.

The observed along channel currents, taken as the principal component of the horizontal velocity, are presented in Figs. 12(d) and 13(d). The currents reflect the stratification patterns. Flood currents occupy most of the water column, while ebbs are concentrated in the surface layer. Strong shear occurs near the bottom layer during ebbs. Bidirectional flow patterns are visible in the OC1 data set during early floods, and lesser floods in the OC2 data set.

As in the case of the AUV data, the model reproduces the salinity field better in low flow than high flow conditions. In Figs. 12(c) and 13(c) the maximum salinity of the model is close to the observed, but the sharp halocline is not captured. In addition, salt retention near the bed is underestimated in magnitude. In Fig. 13(c) the water column becomes entirely fresh during major ebbs in contrast to the observations.

The simulated along channel currents are well in phase with the observations (Figs. 12e and 13e). Phase difference was estimated by fitting a sinusoidal function to the OC2 data points at depth 7.0 ± 0.2 m below the surface. Based on this metric the model is leading the observations by 490 s; In the case of the OC1 data set the model is lagging observations by 370 s. Currents are overestimated by roughly 20%, the difference being most pronounced during ebbs.

3.5.1. Gradient Richardson number

To examine the dynamic stability of the water column, the gradient Richardson number, Ri, was estimated from salinity, temperature and velocity fields,

$$Ri = \frac{N^2}{M^2},$$

$$N^2 = -\frac{g}{\rho_0}\frac{\partial \rho}{\partial z},$$

$$M^2 = \left(\frac{\partial u}{\partial z}\right)^2,$$

where N and M are the buoyancy and vertical shear frequency, respectively.

N and M were estimated by binning the profiler observations on a regular grid, and smoothing the data to reduce noise. Vertical gradients of velocity, temperature and salinity were then evaluated on the grid. Density gradient was obtained by the means of thermal expansion and haline contraction coefficients from the equation of state (Jackett et al., 2006). The same procedure was repeated for modeled fields.
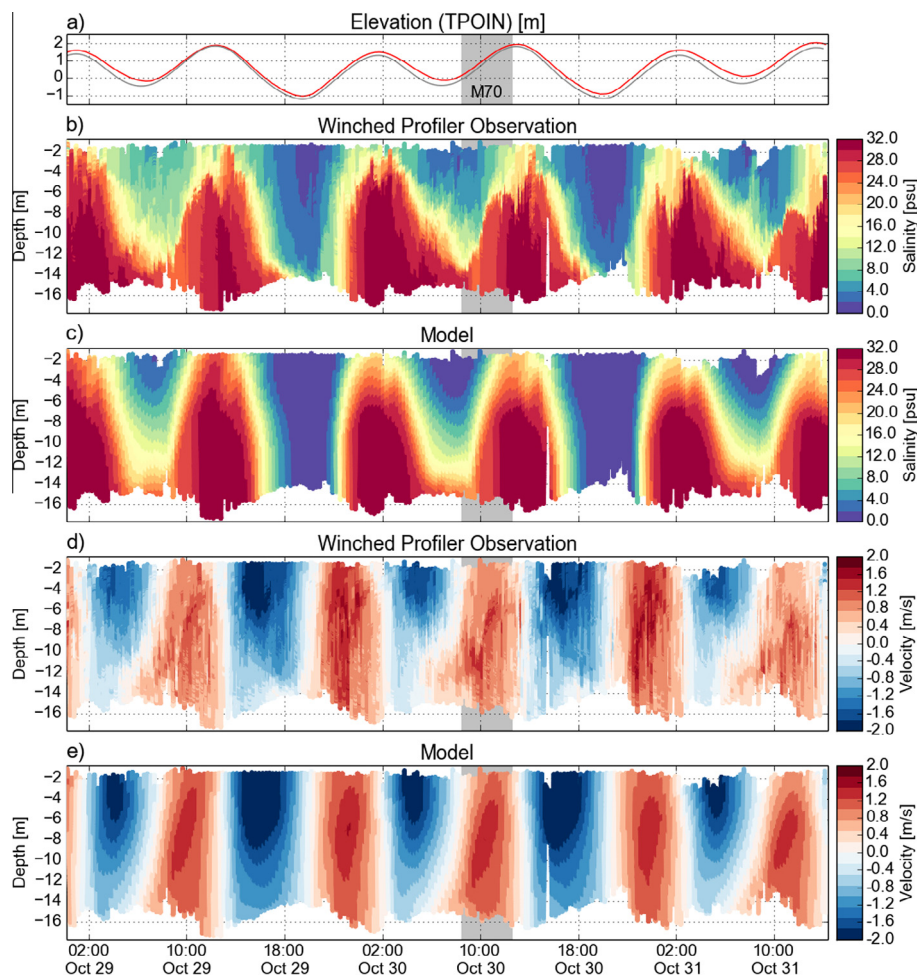
Fig. 13. Winched profiler observations at OC2 versus model output for low flow conditions. (a) water elevation at TPOIN station (red, observed; gray, simulated); (b) observed salinity; (c) modeled salinity; (d) observed along channel velocity; (e) modeled along channel velocity. Along channel velocity in (d) and (e) is taken as the principal component of each data set. The shaded area indicates coinciding AUV mission. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 14 presents M^2, N^2 and Ri for OC1 location (upstream location, weaker tides). The observed shear is largest during ebbs (panel b), and the buoyancy frequency follows the evolution of the salt wedge (panel d). Ri remains well above 0.25 for most of the tidal day and water column, indicating stable stratification (panel f).

Significant mixing ($Ri < 0.25$) only occurs near the bed during floods, related to breakdown of stratification. This is an example of internal tide asymmetry (Jay and Musiak, 1994; Burchard et al., 2004): In the bottom boundary layer, the flood currents are strongest at some distance above the bed (Fig. 12d), which tends to tilt the isopycnals and bring dense water above lighter water mass. The water column becomes unstably stratified and is quickly mixed by intense turbulence, resulting in a nearly homogeneous bottom layer. The process is more pronounced during major floods due to the stronger currents. The strong mixing events lead to nearly vertical isohalines in the tip of the salinity intrusion, and thicker, homogeneous bottom layer. During ebbs the process reverses, straining tending to increase stratification and inhibit mixing. In the latter case Ri remains moderate, however, due to the strong shear related to the ebb currents. There are also patches of high mixing in the surface and middle of the water column at major ebbs, related to mixing in the less stratified surface layer.

The model reproduces similar patterns in M, N and Ri, with some notable differences. High shear induced by the ebb currents, and high buoyancy associated with salinity intrusion are both reproduced. Both of these values are underestimated, however, due to the diffused density and velocity fields. The strong mixing events at the bottom during floods are well captured. In contrast to the observations, however, the model produces unstable mixing events also during major ebbs, due to weaker stratification as the bottom layer becomes overly fresh compared to the observations.

In the case of OC2 (downstream location, stronger tides, Fig. 15), the situation is quite similar. The flow is strongly stratified, except for mixing at the bed during floods, and some mixing events in the surface layer during ebbs. The model reproduces the main features of the flow in this case as well, although M and N are again underestimated. The periodic mixing events at the bed are well reproduced. Similar to OC1 data set, there are additional mixing events during major ebbs near the bed. On Oct 30 almost the entire water column becomes unstably stratified because stratification vanishes entirely (Fig. 13c).
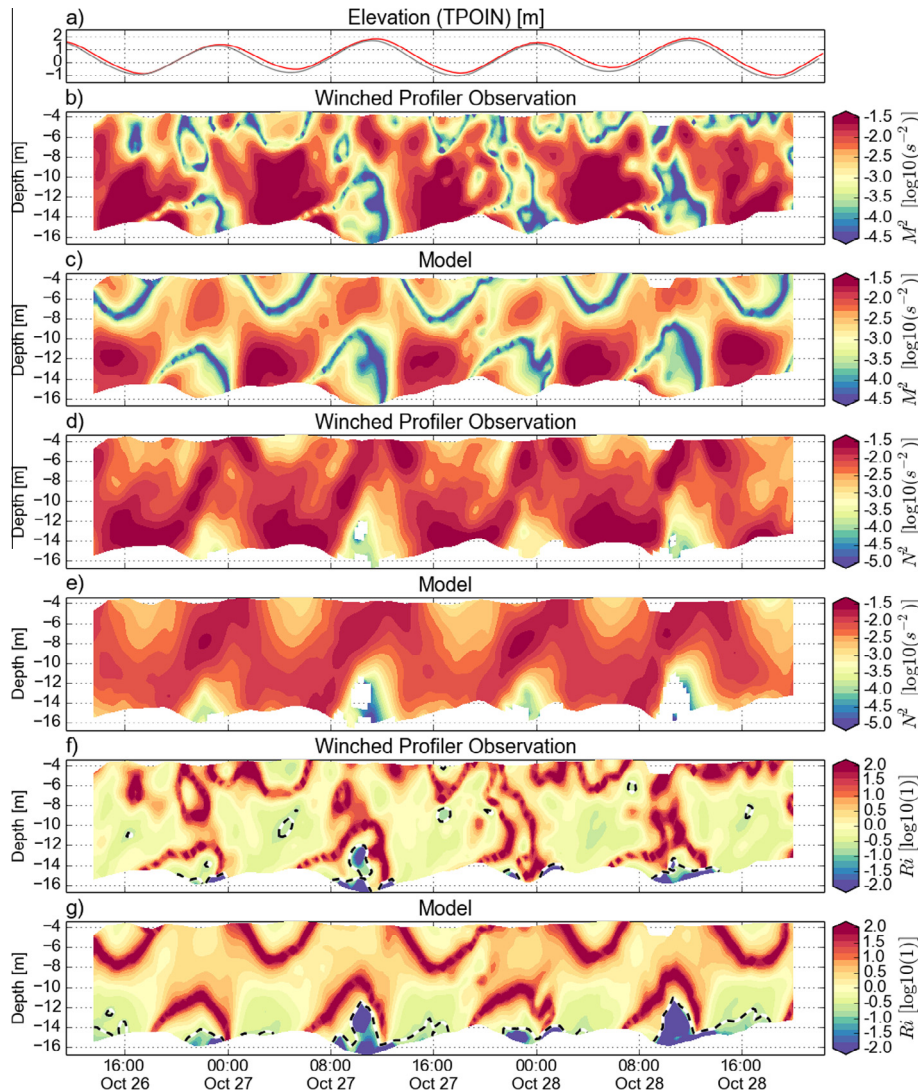

Fig. 14. Shear frequency squared (b and c), buoyancy frequency squared (d and e) and gradient Richardson number (f and g) at OC1 for low flow conditions. The dashed line in (f) and (g) indicates the critical 0.25 contour. Panel (a) shows water elevation at TPOIN station (red, observed; gray, simulated). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5.2. Residual salt transport

Using the fall cruise winched profiler observations, vertical profiles of along-channel velocity, $u = u(z, t)$, and salinity, $S = S(z, t)$, were estimated from the data. The vertical profiles were computed by binning the profiler data to 15 min temporal and 0.5 m vertical resolution. Residual salinity transport was then computed as $\langle uS \rangle$, where $\langle \cdot \rangle$ is the tidal average. Splitting the fields into a tidal average and fluctuating part, $u' = u - \langle u \rangle$ (similarly to Hamilton, 1990), the total salt transport $\langle uS \rangle$ can be expressed as

$$\langle uS \rangle = \langle u \rangle \langle S \rangle + \langle u'S' \rangle.$$

The first term on the right hand side represents advection of mean salinity profile by the residual velocity field. As such it includes transport by classical estuarine circulation. The second term, on the other hand, represents the correlation of tidal signals, and includes tidal pumping. Following Hamilton (1990) these two components are referred to as mean and (dispersive) tidal salt transport, respectively.

Salt transport profiles computed from the two winched profiler data sets are presented in Fig. 16. The temporal average was computed over the last two tidal days in each time series. The observed total transport profiles show a typical two-layer structure with salinity intrusion in the bottom layer. The mean transport component is significant in both cases, particularly for the upstream salt flux near the bed. In the case of the OC1 data set the mean transport dominates, and the tidal effects have only a small influence on the total transport. For the OC2 data set, on the other hand, the tides were stronger and consequently the tidal transport is more important, but still remains small in comparison to the mean transport.

The model underestimates the total landward salt transport in both cases. The mean transport component is significantly different compared to the observations: the modeled mean component tends to carry salt seaward, and the landward transport near the bed is underestimated by at least 75%. The tidal transport, by contrast, is more landward in the model, and partially compensates for the underestimated mean transport. Therefore the salinity intrusion in the model is primarily driven by tidal effects, in contrast to the observations where the mean component is dominant.

These results provide a convenient way to summarize the model skill: Due to numerical mixing the density field is smoother in the model. This leads to underestimated along-estuary density gradient and hence underestimated baroclinic pressure gradient near the bed. Consequently the mean salt transport becomes weaker,
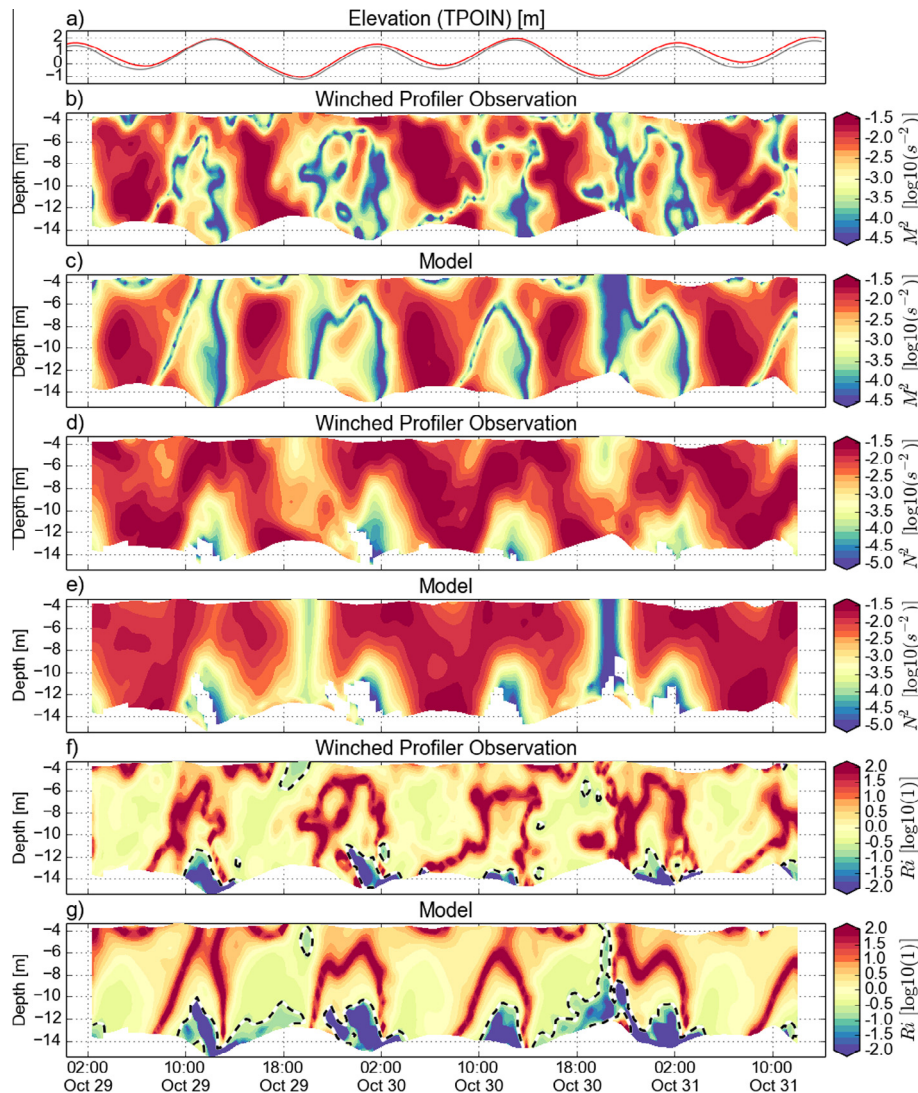
Fig. 15. Shear frequency squared (b and c), buoyancy frequency squared (d and e) and gradient Richardson number (f and g) at OC2 for low flow conditions. The dashed line in (f) and (g) indicates the critical 0.25 contour. Panel (a) shows water elevation at TPOIN station (red, observed; gray, simulated). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
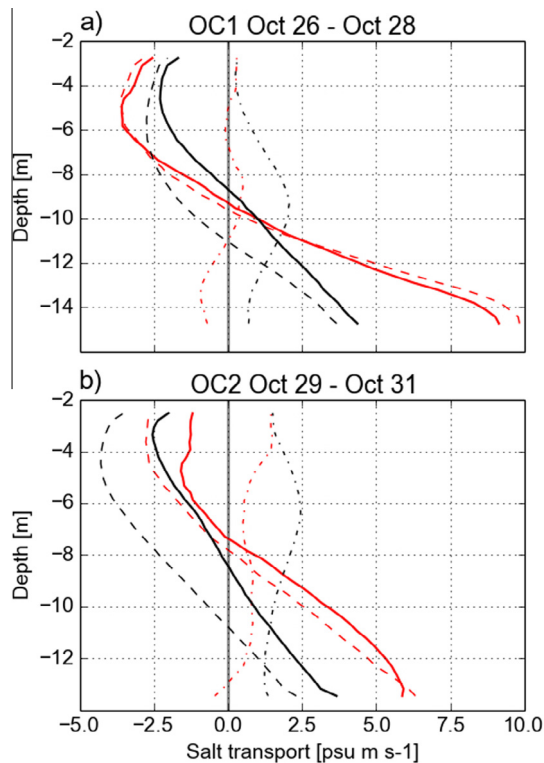
Fig. 16. Residual salt transport profiles for the (a) OC1 and (b) OC2 winched profiler data sets. Red line, observations; Black line, model. Positive values indicate landward transport. The observed total transport $\langle uS \rangle$ (solid red line) shows stronger salt intrusion in the bottom layer than what the model produces (solid black line). The mean ($\langle u \rangle \langle S \rangle$, dashed line) and tidal ($\langle u'S' \rangle$, dash-dotted line) components are qualitatively different from the observations: In the model the mean component tends to carry salt seaward, while the tidal component brings salt upstream. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
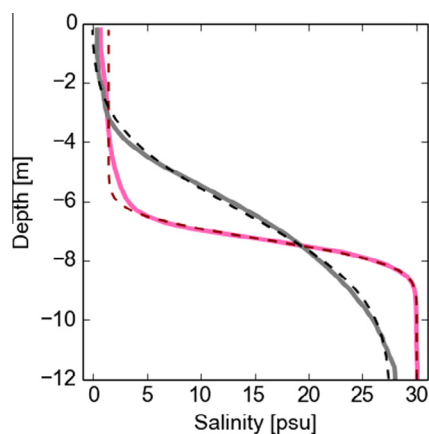
Fig. B.17. Mean salinity profiles estimated from AUV Mission 44 data (solid lines) and the fitted analytical solutions (dashed lines). Pink line, observations; Gray line, model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reducing salinity intrusion in general. In addition, barotropic pressure gradient dominates over baroclinic pressure gradient near the bed leading to underestimated salt retention near the bed during ebbs.

In order to compensate for the weak salinity intrusion, bottom friction needs to be set to a low value in the model, which improves tidal advection of salinity. This delivers more salt in the

estuary during floods, but cannot recover the lost baroclinic nature of the flow; salinity dynamics become predominantly tidally-driven. This discrepancy becomes most evident under high flow conditions when stratification is strongest.

4. Discussion

The AUV and winched profiler observations reflect the known dynamics of the Columbia River estuary: Stratification is strongest under high river discharge and neap tide conditions. During spring tides stratification decreases due to stronger tidal mixing. High flow rates associated with the spring freshet reduce salinity intrusion, pushing the salt wedge out of the North Channel at ebbs.

The state of the estuary is markedly different under low flow conditions. The water column remains stratified over the entire tidal cycle, featuring a smooth vertical density gradient. The stratification is, however, weaker than in high flow conditions, although a sharp halocline may still be observed during floods. Due to the smaller freshwater flux, salinity intrusion is stronger, and there is significant retention of salt in the bottom layer.

The performance of the presented model can be summarized as follows. The timing of salinity intrusion is simulated accurately. The phase difference is within 15 and 8 min in the high and low flow periods, respectively. Salinity tends to be underestimated, although overestimation may also occur at times. Currents are overestimated, by roughly 20% under low flow conditions.

The most salient shortcoming, however, is the model's inability to reproduce sharp density gradients due to numerical mixing. Overly diffused density gradients lead to insufficient decoupling of the surface and bottom layers, and underestimated mean salt transport. In the model landward salt transport is indeed predominantly driven by tidal effects, in contrast to the observations where gravitational circulation dominates.

Sharp density gradients are in general difficult to capture in fixed-grid models. Hofmeister et al. (2010) showed that numerical mixing associated with fixed grids can severely hamper the representation of density-driven flows over fluctuating bathymetry, while using an pycnocline-tracking adaptive vertical grid can substantially reduce spurious mixing. In their examples, realistic density driven overflows across multiple basins (mimicking the bathymetry of the Baltic Sea) was only achieved with adaptive vertical grids. In case of SELFE, high diffusivity is likely related to low accuracy of the time integration scheme, particularly the ELM advection of momentum, or the transport of tracers. SELFE does have a second-order total variation diminishing (TVD) tracer transport scheme, which was not utilized in this study due to its high computational cost. Our shorter calibration runs, however, indicated that the TVD scheme does not yield substantial improvement in terms of the bulk diffusivity of tracers. One could therefore speculate that the issue is related to the vertical grid or the ELM scheme, but a detailed numerical analysis of the exact cause remains a topic for future research.

5. Conclusions

The winched profiler and AUV data sets provide a detailed view of the salt dynamics in the North Channel, and permit model skill assessment on a level that is not often possible. This paper is our first attempt using this data set to better understand the challenges in modeling the Columbia River estuary.

We have shown that the unstructured grid model SELFE is able to simulate the salt dynamics in the Columbia River estuary. The model skill is sensitive to both temporal and spatial resolution, and (to lesser extent) bottom friction, which must be tuned for each application. The model captures the magnitude and phase

of the tidal salinity intrusion with good accuracy for both high and low river discharge conditions. Numerical mixing, however, limits the ability of the model to represent the observed sharp halocline leading to underestimated gravitational circulation and weak residual salt transport, which becomes more apparent under high flow and neap tide conditions. In the future, the spatial and temporal discretization need to be revised in order to reduce numerical mixing. The presented data set provides a rigorous benchmark for this task.

Appendix A. Sensitivity to bottom friction

Skill metrics for different values of bottom roughness length are shown in Table A.3. The station time series error, largely dominated by the SATURN-01 data set, favors small z_0 values, while the AUV and winched profiler data sets yield the best skill with a z_0 roughly a magnitude larger. If SATURN-01 is omitted from the comparison, the station errors are in good agreement with the AUV and winched profiler errors. Reducing bottom friction tends to increase tidal excursion of salinity and shifts the arrival of the salt wedge earlier in the flood tides. Consequently smaller friction value increases the mean salinity at SATURN-01, hence leading to smaller error.

Appendix B. Fitting analytical solution to AUV data

Consider a 1D diffusion equation, defined on the real line $z \in (-\infty, \infty)$,

$$\frac{\partial S}{\partial t} = D\frac{\partial^2 S}{\partial z^2}, \tag{B.1}$$

where D is a constant diffusivity. Assume that the initial condition is a step function,

$$S_0 = \begin{cases} S_{sur}, & \text{if } z \geq z_T, \\ S_{bot}, & \text{otherwise}, \end{cases} \tag{B.2}$$

where S_{sur} and S_{bot} are the surface and bottom layer salinities, respectively, and z_T is the transition depth. In this case the diffusion equation attains a simple analytical solution

$$S(t) = \frac{S_{max} + S_{min}}{2} - \frac{S_{max} - S_{min}}{2}\,\mathrm{erf}\left(\frac{z - z_T}{\sqrt{4Dt}}\right), \tag{B.3}$$

where $\mathrm{erf}(\cdot)$ is the Gauss error function. This solution is valid in a finite domain as well, as long as the front remains far from the surface and bottom boundaries.

Mean vertical salinity profile was estimated from the AUV 44 data set, for time period between 10:27 and 10:48 PST (the first 2.5 km in Fig. 7h), where the salt distribution is fairly stationary. The analytical solution was then fitted to these profiles, using a least-square cost function.

The profiles and fitted functions are presented in Fig. B.17. The analytical solution is in good agreement with the observed profile in the high end of the salinity, but does not match the slope near 2 psu. The observed slope in the surface layer probably indicates residual or advected salinity that the simple 1D model does not capture. The fitted parameter Dt, diffusivity multiplied by a time scale, was 0.90 m² for the observation and 9.80 m² for the model data. Assuming that the salt wedge has been diffused over 6 h period, those translate to diffusivity 4.2×10^{-5} m² s⁻¹ and 4.5×10^{-4} m² s⁻¹, respectively.

References

Baptista, A.M., 1987. Solution of advection-dominated transport by Eulerian–Lagrangian methods using the backwards method of characteristics (Ph.D. thesis). Massachusetts Institute of Technology.

Baptista, A.M., Wilkin, M., Turner, P., Mcc, C., Das, S., Sommerfield, W., Nangia, N., Jay, D., Long, D., Pu, C., Hunt, J., Yang, Z., Darl, J., Farrenkopf, A., 1998. Towards a Multi-Purpose Forecast System for the Columbia River Estuary. Ocean Community Conference '98, Baltimore, MD.

Baptista, A.M., Zhang, Y., Chawla, A., Zulauf, M., Seaton, C., Myers, E.P.I., Kindle, J., Wilkin, M., Burla, M., Turner, P.J., 2005. A cross-scale model for 3D baroclinic circulation in estuary–plume–shelf systems: II. Application to the Columbia River. Cont. Shelf Res. 25, 935–972.

Barron, C.N., Kara, A.B., Martin, P.J., Rhodes, R.C., Smedstad, L.F., 2006. Formulation, implementation and examination of vertical coordinate choices in the global navy coastal ocean model (NCOM). Ocean Modell. 11, 347–375.

Beck, B.C., Baptista, A.M., 1997. WET2: An Eulerian–Lagrangian shallow water FEM model. In: Yeh, H., Liu, P., Synolakis, C. (Eds.), Long-wave runup models. World Scientific Publishing Co., Pte. Ltd, pp. 265–271.

Blumberg, A.F., Mellor, G.L., 1987. A description of a three-dimensional coastal ocean model. In: Heaps, N.S. (Ed.), Three Dimensional Coastal Ocean Models. American Geophysical Union, pp. 1–16.

Burchard, H., Bolding, K., Villarreal, M.R., 1999. GOTM, a general ocean turbulence model. Theory, implementation and test cases. Technical Report EUR 18745 European Commission.

Burchard, H., Bolding, K., Villarreal, M., 2004. Three-dimensional modelling of estuarine turbidity maxima in a tidal estuary. Ocean Dyn. 54, 250–265.

Burla, M., Baptista, A.M., Zhang, Y., Frolov, S., 2010. Seasonal and interannual variability of the Columbia River plume: A perspective enabled by multiyear simulation databases. J. Geophys. Res. (Oceans) 115.

Canuto, V.M., Howard, A., Cheng, Y., Dubovikov, M.S., 2001. Ocean turbulence. Part I: one-point closure model – momentum and heat vertical diffusivities. J. Phys. Oceanogr. 31, 1413–1426.

Casulli, V., Walters, R.A., 2000. An unstructured grid, three-dimensional model based on the shallow water equations. Int. J. Numer. Methods Fluids 32, 331–348.

Chawla, A., Jay, D., Baptista, A.M., Wilkin, M., Seaton, C., 2008. Seasonal variability and estuary-shelf interactions in circulation dynamics of a river-dominated estuary. Estuaries Coasts 31, 269–288.

Elias, E.P.L., Gelfenbaum, G., Van der Westhuysen, A.J., 2012. Validation of a coupled wave-flow model in a high-energy setting: the mouth of the Columbia River. J. Geophys. Res.: Oceans 117.

Geyer, W., 2010. Estuarine salinity structure and circulation. In: Valle-Levinson, A. (Ed.), Contemporary issues in estuarine physics chapter 2. Cambridge University Press, pp. 12–26.

Geyer, W.R., Farmer, D.M., 1989. Tide-induced variation of the dynamics of a salt wedge estuary. J. Phys. Oceanogr. 19, 1060–1072.

Geyer, W.R., MacCready, P., 2014. The estuarine circulation. Annu. Rev. Fluid Mech. 46, 175–197.

Haidvogel, D.B., Arango, H.G., Hedstrom, K., Beckmann, A., Malanotte-Rizzoli, P., Shchepetkin, A.F., 2000. Model evaluation experiments in the North Atlantic

Basin: simulations in nonlinear terrain-following coordinates. Dyn. Atmos. Oceans 32, 239–281.

Hamilton, P., 1990. Modelling salinity and circulation for the Columbia River Estuary. Progr. Oceanogr. 25, 113–156.

Hansen, D.V., Rattray, M., 1966. New dimensions in estuary classification. Limnol. Oceanogr. 11, 319–326.

Hickey, B.M., Banas, N.S., 2003. Oceanography of the U.S. Pacific Northwest Coastal Ocean and estuaries with application to coastal ecology. Estuaries 26, 1010–1031.

Hofmeister, R., Burchard, H., Beckers, J.-M., 2010. Non-uniform adaptive vertical grids for 3D numerical ocean models. Ocean Modell. 33, 70–86.

Hughes, F., Rattray, M., 1980. Salt flux and mixing in the Columbia River Estuary. Estuarine Coastal Mar. Sci. 10, 479–493.

Jackett, D.R., McDougall, T.J., Feistel, R., Wright, D.G., Griffies, S.M., 2006. Algorithms for density, potential temperature, conservative temperature, and the freezing temperature of seawater. J. Atmos. Oceanic Technol. 23, 1709–1728.

Jay, D.A., Flinchem, E.P., 1997. Interaction of fluctuating river flow with a barotropic tide: a demonstration of wavelet tidal analysis methods. J. Geophys. Res.: Oceans 102, 5705–5720.

Jay, D.A., Musiak, J.D., 1994. Particle trapping in estuarine tidal flows. J. Geophys. Res.: Oceans 99, 20445–20461.

Jay, D.A., Smith, J.D., 1990. Circulation, density distribution and neap-spring transitions in the Columbia River Estuary. Progr. Oceanogr. 25, 81–112.

Jay, D.A., Smith, J.D., 1990. Residual circulation in shallow estuaries: 1. Highly stratified, narrow estuaries. J. Geophys. Res.: Oceans 95, 711–731.

Jay, D.A., Smith, J.D., 1990. Residual circulation in shallow estuaries: 2. Weakly stratified and partially mixed, narrow estuaries. J. Geophys. Res.: Oceans 95, 733–748.

Kay, D.J., Jay, D.A., 2003. Interfacial mixing in a highly stratified estuary 2. A method of constrained differences approach for the determination of the momentum and mass balances and the energy of mixing. J. Geophys. Res.: Oceans 108.

Lesser, G., Roelvink, J., van Kester, J., Stelling, G., 2004. Development and validation of a three-dimensional morphological model. Coastal Eng. 51, 883–915.

Liu, Y., MacCready, P., Hickey, B.M., Dever, E.P., Kosro, P.M., Banas, N.S., 2009. Evaluation of a coastal ocean circulation model for the Columbia River plume in summer 2004. J. Geophys. Res.: Oceans 114.

Luettich, R.A., Westerink, J.J., Scheffner, N.W., 1992. ADCIRC: an advanced three-dimensional circulation model for shelves coasts and estuaries, Report 1: theory and methodology of ADCIRC-2DDI and ADCIRC-3DL. Technical Report U.S. Army Corps of Engineers.

Lynch, D.R., Ip, J.T., Naimie, C.E., Werner, F.E., 1996. Comprehensive coastal circulation model with application to the Gulf of Maine. Cont. Shelf Res. 16, 875–906.

MacCready, P., Banas, N.S., Hickey, B.M., Dever, E.P., Liu, Y., 2009. A model study of tide- and wind-induced mixing in the Columbia River Estuary and plume. Cont. Shelf Res. 29, 278–291.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 116, 2417–2424.

Myers, E.P., Baptista, A.M., 2001. Inversion for tides in the Eastern North Pacific Ocean. Adv. Water Res. 24, 505–519.

NGDC, 2006. 2-minute gridded global relief data (ETOPO2v2).

NGDC, 2011. U.S. Coastal Relief Model.

Ralston, D.K., Geyer, W.R., Lerczak, J.A., 2010. Structure, variability, and salt flux in a strongly forced salt wedge estuary. J. Geophys. Res.: Oceans 115.

Sanford, T.B., McNeil, C., Shcherbina, A.Y., Litchendorf, T.M., Kärnä, T., Lopez, J.E., Baptista, A., 2015. CMOP AUV-WP 2012 Columbia River Benchmark data set. Available at <http://dx.doi.org/10.5281/zenodo.13782>.

Song, Y., Haidvogel, D., 1994. A semi-implicit ocean circulation model using a generalized topography-following coordinate system. J. Comput. Phys. 115, 228–244.

Stacey, M.T., Ralston, D.K., 2005. The scaling and structure of the estuarine bottom boundary layer. J. Phys. Oceanogr. 35, 55–71.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G., Roskies, R., Scott, J., Wilkens-Diehr, N., 2014. XSEDE: accelerating scientific discovery. Comput. Sci. Eng. 16, 62–74.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194.

Zhang, Y., Baptista, A.M., 2008. SELFE: A semi-implicit Eulerian–Lagrangian finite-element model for cross-scale ocean circulation. Ocean Modell. 21, 71–96.

Zhang, Y., Baptista, A.M., Myers, E.P., 2004. A cross-scale model for 3D baroclinic circulation in estuary–plume–shelf systems: I. Formulation and skill assessment. Cont. Shelf Res. 24, 2187–2214.