Appendix D. Observed Water Quality Databases

Data Sources

Marine water quality field databases were compiled for the years 2006, 2008, and 2014. Data sources for field databases include Ecology's Marine Monitoring Unit (Ecology MMU), National Oceanographic and Atmospheric Administration (NOAA), University of Washington (UW), and King County. Data collection frequency, duration, and spatial coverage varied between sources depending on the method of collection, which includes CTD (conductivity, temperature, and depth) cast, long term monitoring buoy, or grab sample (Table D1).

Table D1. Field data collection strategies for different sources. CTD = conductivity, temperature, and depth; DO = dissolved oxygen; Chla = chlorophyll; DIC = dissolved inorganic carbon; PAR = photosynthetically active radiation; PMEL = Pacific Marine Environmental Laboratory; TA = total alkalinity; TN = total nitrogen.

Data Source	Collection Approach	Data Type	Parameters	Date/Time Format	2006 Data?	2008 Data?	2014 Data?	Collection Frequency
Ecology MMU	CTD casts	Continuous, profiles	Temperature, Salinity, DO, pH	UTC	Yes	Yes	Yes	2006:~M[3] 2008: M[3] 2014:~M[3]
Ecology MMU	Grab samples	Lab data	NO3, NO2, NH4, Chla, Winkler DO, Salinity	UTC	Yes	Yes	Yes	2006:~M[3] 2008: M[3] 2014:~M[3]
King County	CTD casts	Continuous, profiles	Temperature, PAR, DO, Chla, Salinity	PST/PDT	Yes	Yes	Yes	2006: M[3] 2008: M[3] 2014:~M[3]
King County	Grab samples	Lab data	DO, NH4, NO2/NO3,TN, Chla	PST/PDT	Yes	Yes	Yes	2006: M[3] 2008: M[3] 2014: M[3]
King County	Mooring	Continuous, time series	DO, Chla, Salinity, Temperature, pH (calculated)	PST/PDT	No	Yes	Yes	
NOAA/ PMEL	Buoy	Continuous, profiles/time series	DO, Chla, Salinity, Temperature	PST	Yes	Yes	Yes	
NOAA/UW	Cruises	Continuous, profiles, transects	Temperature, Salinity, DO, Chla,[1] DIC,[2] TA,[2] pH[2] (calculated)	GMT	Yes	Yes	Yes	2006: Twice/year 2008: Twice/year 2014: Four/year

1. Collected in 2006 only
2. Collected in 2008 only
3. Monthly

Data Quality Assurance/Quality Control

All data used in the observed databases were required to meet the acceptance criteria for quality of data as outlined in the Quality Assurance Project Plan (QAPP) for the Salish Sea Model (SSM) (McCarthy et al., 2018). To meet quality assurance project objectives, data used for model calibration required quality assurance/quality control (QA/QC) data quality flags. Data that failed QA screening were not included in any of the observed databases. Further, any data that passed QA screening but contained either (1) outliers, or (2) questionable values due to reporting detection limits, were omitted or retained based on professional judgement (Table D2). With the exception of 2006 PRISM cruises TN196 and TN202, all data contained QA/QC flags. Communication with Jan Newton, Chief Scientist for both of the cruises, verified that QA procedures were undertaken and that any questionable or erroneous data was omitted. As a result of this communication, data from both of the cruises were retained in the 2006 observed database.

Table D2. QA/QC for observed data.

Agency	**Collection Approach**	**Removed Data QA Flags**	**Description**
King County	Grab samples	<MDL	Below mean detection limit
		E	Sample above calibration range
		H	Holding time criteria not met
		SH	Sample handling criteria not met
		R	Result not scientifically defensible based on professional judgement
King County	CTD casts	E	Sample above calibration range
		R	Result not scientifically defensible based on professional judgement
King County	Mooring	All except 111 and 112	QA codes can be found here: https://www.kingcounty.gov/services/environment/water-and-land/puget-sound-marine/marine-mooring/data%20quality%20control.aspx
Ecology MMU	CTD/Lab	All except 2_0_3	First digit 2= Pass Second digit 0 = No specification given Third digit 3= Finalized review
NOAA/UW	Cruises	3	Questionable data
		4	Bad data

Preprocessing

Observed data required preprocessing to allow for a direct comparison between modeled and observed data. Preprocessing consisted of transforming the following observed data attributes to analogous values in the SSM: date/times, water quality parameter units, sample locations, and sample depths.

- Date/time formats for observed data as shown in Table D1 varied between agencies. Using the functions force_tz and ymd_hms from the R Statistical Software package lubridate, date/times were transformed into PST and rounded to the nearest hour to match the SSM (Grolemund, 2011).

- Units of observed water quality parameters (Table D1) were converted to match the model using appropriate factors.

Station/Node Intersection

The SSM uses an unstructured grid system consisting of 16,012 nodes and 25,019 triangular grid elements. For a specific node, hydrodynamic scalar variables (Temperature, Surface Elevation, and Salinity) are calculated by solving governing equations, and fluxes are integrated across the corresponding tracer control element (TCE) (Khangaonkar et al., 2017). A TCE for a given node is composed of connections from the centroids of each triangular element to the midpoint of their edges (Khangaonkar et al., 2017).

In order to relate observed data stations to model nodes, the R sf package was used to assign nodes to stations intersecting a given TCE. To do this, we needed to make the assumption that monitoring station locations were fixed in time. Unfortunately, with the exception of moorings, spatial variability of monitoring locations was observed in certain cases for a given CTD, grab sample, or cruise station. For CTD/grab sample locations, spatial variability was generally very low. Therefore, the location of a given station was represented by the mean sampling location. This was accomplished using R 3.4.3 dplyr package functions: group_by, and summarize. For cruise monitoring stations, sampling locations were not always consistent over time. If a given station was found to be located in more than one TCE, it was considered to be a different station, and dummy variables “a,” “b,” “c,” etc., were affixed to its name. Otherwise, spatial averages were taken as previously mentioned. Station/node intersections were performed using the R 3.4.3 st_join function from the sf package, which creates a list of stations and corresponding nodes.

Observed Data Vertical Layer Binning

The vertical plane of the SSM utilizes a sigma-stretched coordinate system with ten terrain following sigma layers that are distributed with greater layer density near the surface to account for stratification. Layer thickness in the SSM model domain varies with bathymetry, ranging from 0.16 meters in the nearshore top surface layer to 7.6 meters in the top surface layer in deepest areas, and 0.75 meters in nearshore bottom layer to 35 meters in bottom layers of the deepest areas (Pelletier et al., 2017b). The discrete vertical distribution of the SSM model in most circumstances is not directly comparable to observed data profiles without categorically binning sampling depths. Sampling depths for observed data were assigned to model layers ranging from 1 to 10 based on the model layer depth interval in which they fell. For example, at a given node, if a sample depth was between the top and bottom elevations of layer 2, then the sample would be placed into model layer 2. Under certain circumstances, sample depths exceeded the model bottom depth. This typically but not exclusively occurred in nearshore regions where model grid resolution was too coarse to capture steep changes in bathymetry, and these samples were placed into layer 10.

References

Grolemund, G., and H. Wickham. 2011. Dates and times made easy with lubridate. Journal of Statistical Software 40(3): 1–25. http://www.jstatsoft.org/v40/i03.

Khangaonkar, T., W. Long, and W. Xu. 2017. Assessment of circulation and inter-basin transport in the Salish Sea including Johnstone Strait and Discovery Islands pathways. Ocean Modelling 109: 11–32.

McCarthy, S., C. Figueroa-Kaminsky, A. Ahmed, T. Mohamedali, and G. Pelletier. 2018. Quality Assurance Project Plan: Salish Sea Model Applications. Publication 18-03-111. Washington State Department of Ecology, Olympia. https://fortress.wa.gov/ecy/publications/SummaryPages/1803111.html.

Pelletier, G., L. Bianucci, W. Long, T. Khangaonkar, T. Mohamedali, A. Ahmed, and C. Figueroa-Kaminsky. 2017b. Salish Sea Model Ocean Acidification Module and the Response to Regional Anthropogenic Nutrient Sources. Publication 17-03-009. Washington State Department of Ecology, Olympia. https://fortress.wa.gov/ecy/publications/SummaryPages/1703009.html.