Hydrodynamic controls on oxygen dynamics in a riverine salt-wedge estuary, the Yarra River estuary, Australia.

Louise C. Bruce[1,2,*], Perran L. M. Cook[3], Ian Teakle[4], Mathew R. Hipsey[1,2]

Supplement – Model Evaluation

S.1 Methods

Water elevation measurements from four stations in the domain, (Spencer Street, Burnley Depot, Hawthorn and Dight's Falls refer to Figure 1) were compared to model output. For salinity and oxygen, 9 along channel profile transects of surveys conducted between September 2009 and June 2009 were used (refer to Roberts et al., 2012)). For these data comparisons were made for time series data, field profiles, degree of stratification as well as upstream extent of salt wedge, hypoxia and anoxia. For each survey, depth profile data were collected at 5-6 of the 17 stations indicated in Figure 1 (chainage listed in Table A1). Time series comparisons of surface and bottom salinity and oxygen were made at three representative stations where greatest data were available: Morell Bridge (in the down stream end of the estuary); Scotch College (middle of the estuary) and Bridge Road (upstream end of the estuary) (Figure 1).

In this study we evaluate five alternative measures of model performance. The purpose of calculating alternative model fit parameters are two fold: to enable comparison to similar modelling studies with various methods of performance evaluation, and secondly because different methods of model evaluation tell us different things about model performance (Bennett et al., 2013). Measures of model fit were calculated as:

1) Normalised mean absolute error (NMAE; Alewell and Manderscheid, 1998):

$$NMAE = \frac{\sum_{i=1}^{N}(|P_i - O_i|)}{N\bar{O}}$$

(1)

2) Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}}$$

(2)

3) Model Efficiency (MEF; Murphy, 1988; Nash and Sutcliffe, 1970):

$$MEF = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$$

(3)

4) Model Skill Score (MSS; Willmott (1981)):

$$MSS = \frac{\sum_{i=1}^{N}|P_i - O_i|^2}{\sum_{i=1}^{N}|P_i - \bar{O}| + |O_i - \bar{O}|}$$

(4)

5) Correlation coefficient (r)

$$r = \frac{\sum_{i=1}^{N}(P_i - \bar{P})(O_i - \bar{O})}{\left[\sum_{i=1}^{N}(P_i - \bar{P})^2 \sum_{i=1}^{N}(O_i - \bar{O})^2\right]^{1/2}}$$

(5)

where N is the number of observations, O_i and P_i, the "i[th]" observed and model predicted data and $\bar{O}$ and $\bar{P}$ the mean observed and model predicted data respectively.

NMAE is a measure of the absolute deviation of simulated values from observations, normalized to the mean; a value of zero indicated perfect agreement and greater than zero an average fraction of the discrepancy normalized to the mean. Similarly RMSE is a measure of the average square error with values near zero indicating a close match. MEF is a measure of the square of the deviation of simulated values from observations, normalized to the standard deviation of the observed data. For both MEF and MSS, a maximum value of one indicates perfect fit; zero indicates that the model provides equal predictive skill as assuming mean observed data. Ralston et al. (2010) showed MEF to be an unreliable measure of model fit, however we have included this metric as it is commonly used allowing us to compare model fit with similar estuary modelling studies. The correlation coefficient gives an indication of the linear relationship between observed and predicted data and is the most common measure for assessing aquatic models (Ahronditsis and Brett 2004).

S.2 Results

The model performed well in measure of fit for surface water elevation in all four gauged stations (Table A2). All correlation coefficients were > 0.95, the lowest measure of fit occurred at Burnley Depot located just downstream of the first horseshoe bend. Model skill scores were all reasonably high (>0.8) and NMAE reasonably low (<0.3) with the exception of Burnley Depot. Patterns of model fit for water level elevation were consistent for all four measures of skill score.

Simulated time series of the surface and bottom concentrations of salinity and oxygen followed both the seasonal and episodic patterns observed in the field measurements (Figure A1). Comparisons of along channel transects showed good agreement between model and data including the extent of salt-wedge propagation and associated regions of oxygen depletion (Figure A2). Measurements of model fit for salinity and oxygen varied significantly from position in the water column and region (Table A3). Whilst measures of MSS, MEF and r^2 were greater overall for salinity (0.42/0.85/0.72) than oxygen (0.29/0.80/0.65) measures of NMAE were lower for oxygen (0.28) than salinity (0.47). For salinity, measures of model fit indicated that the model predicted bottom salinities better than surface and surface oxygen better than bottom (Table A3). The model simulated salinity of the lower reaches of the estuary better than the upstream end, with negative values of MSS indicating that the model predicted worse than using an observational average for the regions > 10km upstream of Spencer Street Bridge. For simulated oxygen, the patterns of model fit were similar although the model performed better than salinity in the upper reaches of the estuary. In fact, for r^2 the highest value was in the region > 15km from Spencer Street Bridge ($r^2 = 0.74$).

S.3 Discussion

Whist there has been significant progress made in modeling morphologically complex coastal environments (Chen et al., 2003, 2007; Liu et al., 2007; Luyten et al., 2003; Xu et al., 2012) simulating the finely resolved pycnocline of a salt wedge riverine estuary with tight curvature has remained a challenge (Kurup et al., 2000; Li et al., 2005; Oey et al., 1985; Warner et al., 2005). Progress has been made with respect to horizontal grid co-ordinates able to accurately represent more complex estuarine bathymetries, including a move from Cartesian (Simanjuntak et al., 2011) to curvilinear (Burchard et al., 2004) and finite element models (Wang and Justić, 2009). The recent development of the finite volume method that combines finite element grid structures with finite-difference methods has led to improved computational efficiency in estuarine modeling studies (Chen et al., 2003; Wang and Justić, 2009). Most models use sigma-coordinates in the vertical that may be less suited to resolve sharp pycnoclines in a riverine estuary due to problems with numerical diffusion. Depending on the bathymetric complexity of a particular system, the choice of physical model is crucial to the correct simulation of the salt-wedge dynamics and therefore critical in determining the extent of oxygen depletion. After attempting to model the Yarra River estuary with both cartesian and curvilinear grid structures, the choice of flexible mesh structure gave the greatest accuracy in predicting the position and extent of the salt wedge dynamics. Where the model fit results were low, ideas for model improvement are discussed below.

Whilst rigorous analysis of the accuracy or “model fit” of numerical models is considered crucial, particularly when a model is used for predictive management decisions (Stow et al., 2009), quantitative measurements of model fit are rarely presented (Arhonditsis and Brett, 2004; Mooij et al., 2010). A further challenge in the comparison of model performance is a lack of consistent model fit metrics reported. In this study we included five alternative measures of model fit in order to compare against similar estuarine modelling studies. The measures for surface water elevations were generally high and in close agreement with those measured by Ralston et al. (2010 MSS=0.78-0.92 r=0.99), Xu et al. (2012 RMSE = 0.13 MSS=0.8 MEF=0.9 r=0.91) and Warner et al. (2005 MSS=0.85-0.95). The greatest deviation from observed data occurred at Burnley Depot (13.1km) and can be attributed to the over simplification of bathymetry in the mesh cells just downstream of the sampling station. In calculating the measures of model fit for surface elevations, the number of sampling points in the time series record (N=17524) was relatively high which could account for the observed consistency in model fit measures since a large sample space minimises bias and sensitivity to phase errors. In general, the high measures of model fit for surface elevation were an indication that the mesh bathymetry, inflows and tidal boundary forcing were well represented in the model. An improved mesh structure with higher resolution particularly of the river banks and regions of high curvature could lead to improved prediction of water elevations in the estuary.

Much progress has been made in the accurate prediction of stratification in the lower reaches of estuaries (Ralston et al., 2010) with less success in the upper reaches of narrow, shallow estuaries (Sharples et al., 1994). In general for this application, measures of model fit for salinity decreased with upstream distance. Values of MSS and r^2 compared well to Ralston et al. (2010) for bottom salinity (0.70-0.94/0.88-0.99) and surface salinity (0.29-0.86/0.77-0.94), Xu et al. (2012) for overall salinity (0.88/0.94) and Warner et al. (2005) for salinity (MSS=0.85). The slope and offset calculations suggest that the model is under-predicting lower salinities and over predicting higher salinities indicating the degree of vertical mixing was under-predicted. Some deviations from observed data could be attributed to uncertainty in the boundary conditions, particularly the meteorological data and absence of smaller tributary inflows such as storm water drains. Calibration of mixing parameters and a higher resolution bathymetry may also lead to improved (less stratified) predictions of salinity distribution.

Measures of model fit for oxygen were generally lower than for salinity and surface water elevations. This difference is consistent with a general trend in modelling of aquatic ecosystems (Arhonditsis and Brett, 2004; Burchard et al., 2006) where model fit measures decrease from the physics, through nutrients to higher order plankton (Allen et al., 2007a). Whilst salinity is dependent only on the hydrodynamic model equations, oxygen concentrations may additionally include error related to biotic factors represented in the model, which are often less controlled and not as easily parameterised. It is interesting to note that for NMAE the model performed better for oxygen than salinity. Since NMAE is a measure of the absolute error, normalized to the mean the difference in model performance is attributed to the greater relative spread of data for oxygen (standard deviation = 94% of mean) compared to salinity (standard deviation = 49% of mean). These results emphasise the need to calculate a number of measures of model to fit to evaluate model performance against variables with a range of variance.

For salinity, higher values of model fit for the bottom layer compared to the surface layer are indicative of a better representation of the salt-wedge propagation and less on mixing in the surface layers. On the other hand for oxygen, higher values of model fit for the surface layer compared to the bottom reflect the dominance of air-water surface exchange at the top, and highlight a need to focus calibration methods on improved parameterisation of sediment oxygen demand. Simulated outputs indicated significant diurnal variations in the position and extent of both the salt wedge and oxygen depleted waters similar to those observed by data loggers employed in 2008 outside the simulated period (Roberts et al., 2012). Whilst a simplified mesh of just a few hundred cells gave acceptable levels of model fit related to the propagation and strength of the salt wedge intrusion, a mesh with increased resolution in the cross section would allow improved predictability of the fine resolution of the pcynocline and lateral mixing.

For prediction of oxygen concentrations, closeness of model fit were limited by the error associated with parameter estimations and the assumption of homogeneous sediment oxygen demand. Whilst a best fit Arrhenius oxygen limitation was used to parameterize the oxygen sediment flux parameters using measured data from laboratory studies, there was a substantial amount of scatter in the data. This scatter indicates that other limitation factors were at play including sediment organic matter and that the sediment flux is also dependent on sediment type and therefore location in the estuary. Errors associated with the experimental method also contribute to uncertainty in estimation of model parameters and predicted concentrations. Furthermore errors associated with boundary condition interpolation calculations as discussed in the methods contributed to errors associated with model prediction.

Each model skill equation used to measure the performance of the model results gave slightly different indications of model fit which is why a number of measures is preferable as they tell us different things about model performance (Bennett et al., 2013; Stow et al., 2009). Traditional methods of fit such as r^2, NMAE and RMSE were useful in determining absolute deviations from observed and calculated for each spatial domain in the vertical and horizontal helped direct focus on further model refinement. Measures of model fit that relates error to variability in observational error, such as MEF and MSS, on the other hand are useful measures of model efficiency that can be compared to similar modelling studies (Allen et al., 2007b; Stow et al., 2009). For this study we found that in regions with low variance in observed data such as the upper reaches of the estuary values of MEF and MSS gave little benefit in evaluation of model performance. On the other hand in the middle regions with more dynamic patterns of salinity and oxygen depletion, MEF and MSS indicated a strong model performance. The different measures of model performance demonstrated that whilst the general dynamics of the shifting salt-wedge and associated anoxia were well reproduced by the model, refinement of boundary conditions and mesh morphometry is required to improve model performance near the domain boundaries.

References

Alewell, C. and Manderscheid, B.: Use of objective criteria for the assessment of biogeochemical ecosystem models, Ecol. Modell., 107(2-3), 213–224, doi:10.1016/S0304-3800(97)00218-4, 1998.

Allen, J. I., Holt, J. T., Blackford, J. and Proctor, R.: Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM, J. Mar. Syst., 68(3-4), 381–404, doi:10.1016/j.jmarsys.2007.01.005, 2007a.

Allen, J. I., Somerfield, P. J. and Gilbert, F. J.: Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models, J. Mar. Syst., 64(1-4), 3–14, doi:10.1016/j.jmarsys.2006.02.010, 2007b.

Arhonditsis, G. and Brett, M.: Evaluation of the current state of mechanistic aquatic biogeochemical modeling, Mar. Ecol. Prog. Ser., 271, 13–26, doi:10.3354/meps271013, 2004.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. a., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. a., Robson, B., Seppelt, R., Voinov, A. a., Fath, B. D. and Andreassian, V.: Characterising performance of environmental models, Environ. Model. Softw., 40, 1–20, doi:10.1016/j.envsoft.2012.09.011, 2013.

Burchard, H., Bolding, K., Kühn, W., Meister, A., Neumann, T. and Umlauf, L.: Description of a flexible and extendable physical–biogeochemical model system for the water column, J. Mar. Syst., 61(3-4), 180–211, doi:10.1016/j.jmarsys.2005.04.011, 2006.

Burchard, H., Bolding, K. and Villarreal, M. R.: Three-dimensional modelling of estuarine turbidity maxima in a tidal estuary, Ocean Dyn., 54(2), 250–265, doi:10.1007/s10236-003-0073-4, 2004.

Chen, C., Liu, H. and Beardsley, R.: An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: application to coastal ocean and estuaries, J. Atmos. Ocean. Technol., 20(1), 159:186 [online] Available from: http://journals.ametsoc.org/doi/abs/10.1175/1520-0426(2003)020%3C0159%3AAUGFVT%3E2.0.CO%3B2 (Accessed 13 November 2012), 2003.

Chen, X., Zhu, L. and Zhang, H.: Numerical simulation of summer circulation in the East China Sea and its application in estimating the sources of red tides in the Yangtze River estuary and adjacent, J. Hydrodyn. Ser. B, 19(2001), 272–281 [online] Available from: http://www.sciencedirect.com/science/article/pii/S1001605807600596 (Accessed 7 May 2013), 2007.

Kurup, R. G., Hamilton, D. P. and Phillips, R. L.: Comparison of two 2-dimensional, laterally averaged hydrodynamic model applications to the Swan River Estuary, Math. Comput. Simul., 51(6), 627–638, doi:10.1016/S0378-4754(99)00146-9, 2000.

Li, M., Zhong, L. and Boicourt, W. C.: Simulations of Chesapeake Bay estuary: Sensitivity to turbulence mixing parameterizations and comparison with observations, J. Geophys. Res., 110(C12), C12004, doi:10.1029/2004JC002585, 2005.

Liu, W.-C., Chen, W.-B., Cheng, R. T., Hsu, M.-H. and Kuo, A. Y.: Modeling the influence of river discharge on salt intrusion and residual circulation in Danshuei River estuary, Taiwan, Cont. Shelf Res., 27(7), 900–921, doi:10.1016/j.csr.2006.12.005, 2007.

Luyten, P., Jones, J. and Proctor, R.: A numerical study of the long-and short-term temperature variability and thermal circulation in the North Sea, J. Phys. …, 37–56 [online] Available from: http://journals.ametsoc.org/doi/pdf/10.1175/1520-0485(2003)033%3C0037%3AANSOTL%3E2.0.CO%3B2 (Accessed 7 May 2013), 2003.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., Degermendzhy, A. G., DeAngelis, D. L., Senerpont Domis, L. N., Downing, A. S., Elliott, J. A., Fragoso, C. R., Gaedke, U., Genova, S. N., Gulati, R. D., Håkanson, L., Hamilton, D. P., Hipsey, M. R., 't Hoen, J., Hülsmann, S., Los, F. H., Makler-Pick, V., Petzoldt, T., Prokopkin, I. G., Rinke, K., Schep, S. a., Tominaga, K., Dam, A. a., Nes, E. H., Wells, S. a. and Janse, J. H.: Challenges and opportunities for integrating lake ecosystem modelling approaches, Aquat. Ecol., 44(3), 633–667, doi:10.1007/s10452-010-9339-3, 2010.

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

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290 [online] Available from: http://www.sciencedirect.com/science/article/pii/0022169470902556 (Accessed 14 May 2013), 1970.

Oey, L., Mellor, G. and Hires, R.: A Three-Dimensional Simulation of the Hudson-Raritan estuary. Part II: Comparison with Observation, J. Phys. Ocean., 15(12), 1711–1720 [online] Available from: http://www.gfdl.noaa.gov/bibliography/related_files/lyo8502.pdf (Accessed 7 May 2013), 1985.

Ralston, D. K., Geyer, W. R. and Lerczak, J. a.: Structure, variability, and salt flux in a strongly forced salt wedge estuary, J. Geophys. Res., 115(C6), C06005, doi:10.1029/2009JC005806, 2010.

Roberts, K. L., Eate, V. M., Eyre, B. D., Holland, D. P. and Cook, P. L. M.: Hypoxic events stimulate nitrogen recycling in a shallow salt-wedge estuary: The Yarra River Estuary, Australia, Limnol. Oceanogr., 57(5), 1427–1442, doi:10.4319/lo.2012.57.5.1427, 2012.

Sharples, J., Simpson, J. and Brubaker, J.: Observations and Modelling of Periodic Stratification in the Upper York River Estuary, Virginia, Estuarine, Coast. Shelf …, 38, 301–312 [online] Available from: http://www.sciencedirect.com/science/article/pii/S0272771484710213 (Accessed 11 June 2013), 1994.

Simanjuntak, M. A., Imberger, J., Nakayama, K. and Ishikawa, T.: Benthic and interfacial mixing in a strongly-stratified estuary, J. Hydraul. Res., 49(6), 791–798, doi:10.1080/00221686.2011.607304, 2011.

Stow, C. a., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. a. M., Rose, K. a. and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems, J. Mar. Syst., 76(1-2), 4–15, doi:10.1016/j.jmarsys.2008.03.011, 2009.

Wang, L. and Justić, D.: A modeling study of the physical processes affecting the development of seasonal hypoxia over the inner Louisiana-Texas shelf: Circulation and stratification, Cont. Shelf Res., 29(11-12), 1464–1476, doi:10.1016/j.csr.2009.03.014, 2009.

Warner, J. C., Geyer, W. R. and Lerczak, J. A.: Numerical modeling of an estuary: A comprehensive skill assessment, J. Geophys. Res., 110(C5), C05001, doi:10.1029/2004JC002691, 2005.

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

Xu, J., Long, W., Wiggert, J., Lanerolle, L., Brown, C., Murtugudde, R. and Hood, R. R.: Climate forcing and salinity variability in Chesapeake Bay, USA, Estuaries …, 35(1), 237–261, doi:10.1007/s12237-011-9423-5, 2012.

Table S1. Chainage (distance along thalweg from Port Philip Bay) of main sampling stations used for model evaluation.

Station Name	No	Chainage
Spencer Street Bridge	0	7.4
Morell Bridge	4	10.6
Burnley Depot	5	12.6
Scotch College	8	15.0
Hawthorn Reserve	9	15.3
Bridge Road	13	16.4
Dight's Falls	19	21.9

Table S2. – Model fit metrics for surface water elevation. Normalised mean absolute error (NMAE), model efficiency (MEF), model skill score (MSS) and correlation coefficient (r).

Station	NMAE	MEF	MSS	r
SPEN	0.11	0.98	0.99	0.99
BURN	0.55	0.69	0.93	0.96
HAWTH	0.31	0.83	0.96	0.97
COLL	0.20	0.91	0.98	0.96
Total	0.31	0.89	0.97	0.95

Table S3. – Model fit metrics for salinity and oxygen. Normalised mean absolute error (NMAE), root mean square error (RMSE), model efficiency (MEF), model skill score (MSS) and correlation coefficient (r).

Variable	Salinity					Oxygen				
Data	NMAE	RMSE	MEF	MSS	r	NMAE	RMSE	MEF	MSS	r
Surface	0.66	0.42	0.05	0.69	0.50	0.14	0.19	0.21	0.78	0.64
Bottom	0.30	0.13	0.27	0.79	0.64	0.40	0.23	0.24	0.76	0.58
0-5km	0.24	0.35	0.33	0.85	0.73	0.17	0.44	0.41	0.85	0.75
5-10km	0.33	0.44	0.60	0.89	0.80	0.28	0.49	0.49	0.84	0.72
10-15km	0.98	0.43	-0.36	0.69	0.44	0.32	0.21	0.21	0.77	0.62
>15km	1.58	0.95	-0.94	0.64	0.48	0.40	0.14	-0.14	0.75	0.74
Total	0.47	0.31	0.42	0.85	0.72	0.28	0.29	0.29	0.80	0.65

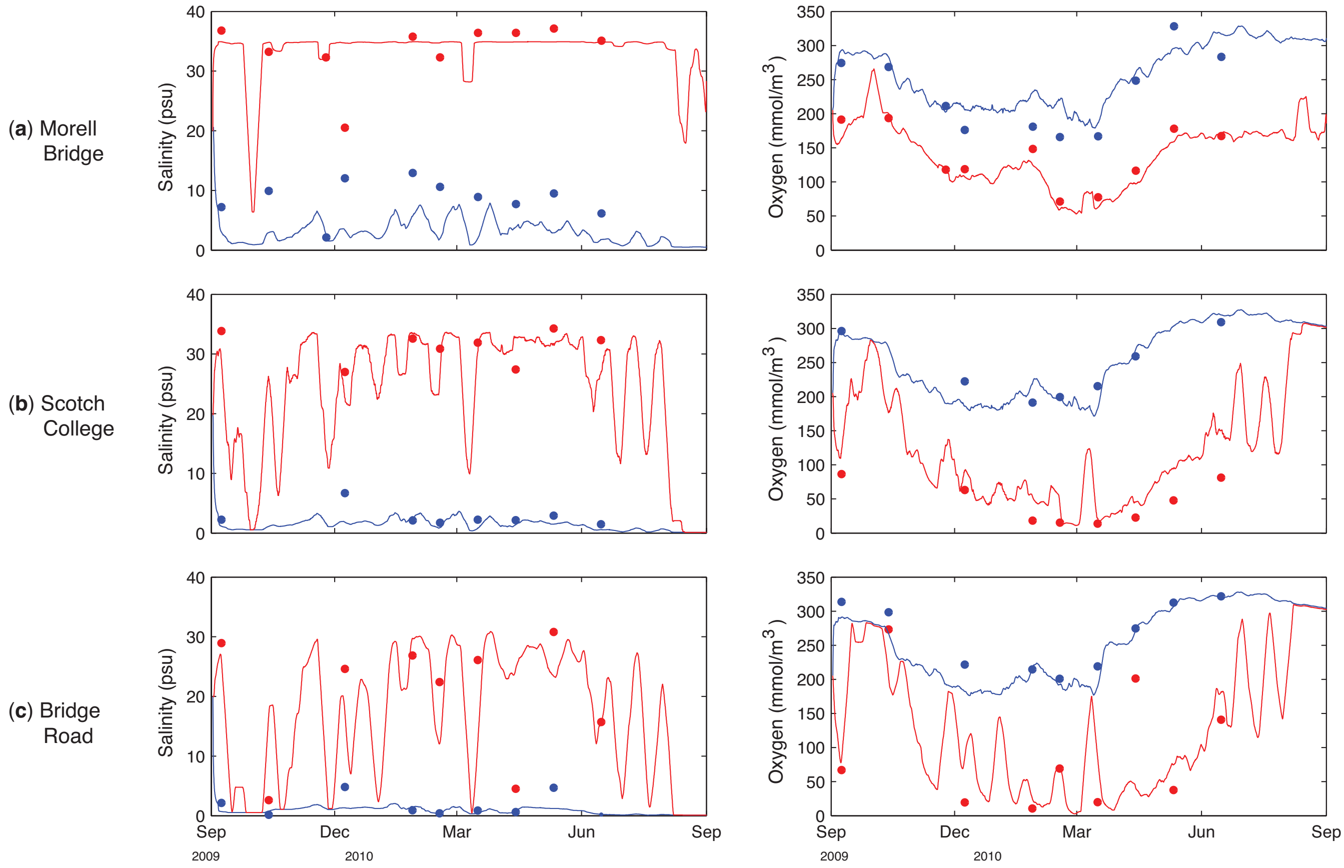

Figure S1. - Time series data for surface (blue) and bottom (red) salinity and oxygen for (a) Morell Bridge (b) Scotch College and (c) Bridge Road. Dots represent observed and solid lines simulated data. Refer to error metrics in Appendix A, supplementary material for measures of model fit.

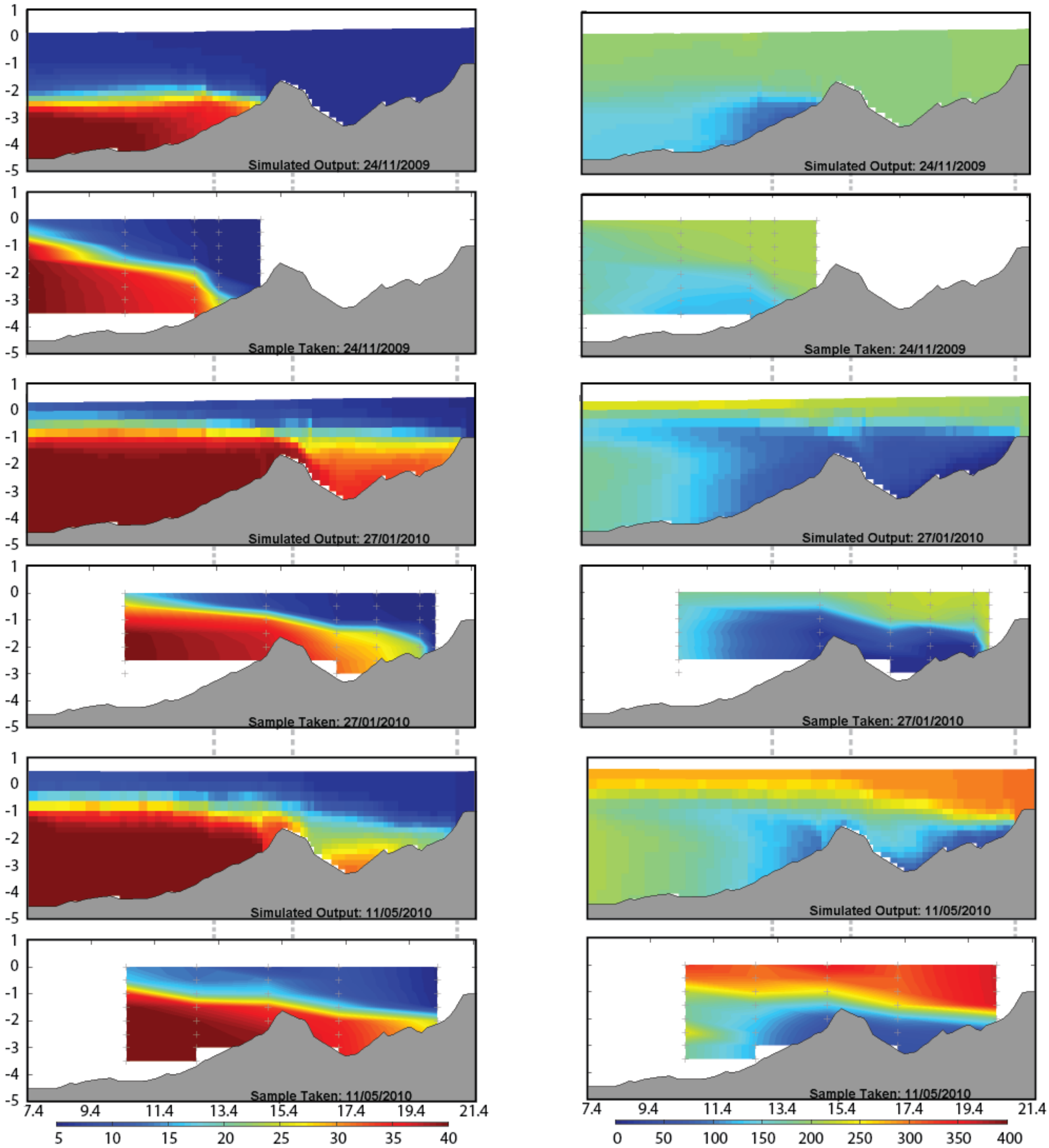

Figure S2. – Along channel transects of salinity (left) and oxygen (right), for three different time periods showing simulated output above field data contoured around available profile data (+).