Biogeosciences Discuss., 12, C9282–C9283, 2016
www.biogeosciences-discuss.net/12/C9282/2016/

Biogeosciences Discussions

Interactive comment on "Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison" *by* I. D. Irby et al.

Anonymous Referee #2

Received and published: 20 January 2016

Overall, this is a well-conceived modeling study that compares predictions of eight coupled hydrodynamic-biogeochemical models that were independently developed for the Chesapeake Bay against the data collected on biweekly to monthly monitoring cruises conducted during 2004-2005. In terms of the number of models involved, this is certainly one of the more comprehensive model comparisons conducted for coastal ecosystems. The members of the team are skillful modelers that have extensively published on the subject and the methods and conclusions are generally sound and scientifically defensible. The paper is well written and suitable for publication, subject to minor revision as suggested below.

The conclusion that all models predict the seasonal dynamics of dissolved oxygen reasonably well, regardless of their structural complexity of spatial resolution, is not surprising. Extensive model comparisons conducted with climate models have thought us a very important modeling lesson – eight climate models that produce nearly identical hindcasts for the past 2,500 years, strongly disagree in their predictions for the next 85 years for the same climate scenario. I guess there is simple answer to that - calibration. Modelers have become very good in calibrating their models, and given sufficient time and data, even a model of dubious mechanistic value will end up displaying a remarkable skill. The only way to critically evaluate the model results would be if they are subjected to a rigorous validation using data to which the models were not exposed during calibration. Because of the different data requirements, this would be very difficult to accomplish with a large number of fairly complex models, and I am not suggesting that the authors embark on that journey. However, some discussion would be needed to clarify whether the 2004-2005 data set that was used for model comparison was also used for model calibration.

My second point is that I would like to have seen a more detailed analysis of the model-data comparison. For example, Fig. 9 shows that models collectively predict a duration of hypoxia compared to the measurements, and that the predicted onset of hypoxia during 2005 lags substantially with respect to the measurements. As much as I appreciate Taylor and target diagrams, I think that simple scatter plots of predicted versus observed DO values for individual models would have been very useful in that regard.

My third point concerns the selection of model data for monthly comparison. I am not sure what the word "monthly" refers to. Were the model results outputted to match the dates of the biweekly to monthly monitoring cruises, or were they averaged for the entire month?

Interactive comment on Biogeosciences Discuss., 12, 20361, 2015.