

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

P. MacCready (Referee)

parker@ocean.washington.edu

Received and published: 5 January 2016

The authors systematically compare the skill of 8 3D models of Chesapeake Bay circulation and biogeochemistry. They focus on hypoxia, but consider other related properties such as mixed layer depth. They find that all models do a reasonable job at simulating hypoxia compared to two years of ~monthly observations at 13 stations. Like temperature (with which the models also have high skill) oxygen has a large seasonal cycle, contributing to its predictability. All models had poor skill at predicting the depth of the start of the hypoxic layer (very important for the ecosystem and management). The authors show that this problem is related to lack of skill predicting the

C8848

density mixed layer depth.

This is an important piece of work. This level of model inter-comparison has rarely or never occurred for estuarine systems. The paper is well-written, and the figures well-chosen. I have only a few smaller comments, and recommend that it be accepted with minor revisions.

Smaller Comments:

Page 20371, lines 7-9. The "minimum stratification criterion" is mentioned here, and it seems like a good way to ignore casts with minimal gradients, but I could not find where this criterion is defined. Please clarify.

Page 20372, Lines 5-6. The phrase about "the skill of a model defined as the mean of the observations" was unclear. In general the authors do a good job explaining the statistical tests, but in this case another sentence might help.

Page 20373, bottom. It might help to explain when and where it is of greatest value (e.g. to managers) to get the DO right, and why.

Page 20374, Line 13. Why is it that all the models have the same biases in the stratification field?

Page 20375, Line 20. It is interesting that the mean of the models has these timing errors, but I'm not sure what it shows. Two years is too few to say anything statistical about timing. Probably OK to mention, though.

Page 20378, Lines 24-25. This sentence is unclear. In what way are the biological drivers "not ... spatially explicit"?

Fig. 1. Give the length scale in km.

Fig. 8b. The very poor correlation of Chl seems to make sense for this inherently patchy process. What is more perplexing is the large underestimate of the standard deviation. Any thoughts?

C8849

Fig. 9. Please make the x-axis ticks more regular so that the same time in different years is easier to compare visually.

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

C8850