This is a well written paper that provides a management strategy evaluation for a stock with seasonal natal homing movement patterns, using assessment models with varying levels of spatial complexity. This paper is suitable for publication after major revisions that address my comments below. Most of my comments require more text to better explain methods and results. A few grammatical suggestions are provided in the attached pdf.

Major Comments

1. The authors need an equation for the basic cohort decay equation to explain more clearly what the NAA REs are.
2. Describe how the NA REs were modelled. IID or 2DAR1? I suspect that IID Res have limited ability to capture model mis-specification because these process errors tend to deviate from zero less than 2DAR1 errors. I suggest how to model NAA REs should be a discussion topic.
3. L74 and L487-488. This depends on what the authors mean by true fishing selectivity, which needs to be clarified in the paper. If there is spatial heterogeneity in the age distribution of the stock and area-specific fleets, then the selectivity of the fleets for the entire stock will differ, although selectivity for the area sub-stocks may be the same. If the trends in stock numbers at age over time are the same in the areas then this approach is OK I think. [my rant] However, if there is something like source-sink dynamics or environmental shifts in distribution then the trends in the areas will not be the same. I suspect that spatial dynamics are a major reason for "conflicts" in stock assessment inputs. The conflict is actually in the “too simple” space-aggregated model and not the inputs.
4. L109. Closed-loop simulations need to be explained a little more, perhaps later in the paper, for those not familiar with MSE (like many graduate students). This will make the paper more accessible to a broader readership
5. L182. CV=0.1 for survey seems like a low value. For example, the confidence intervals in Figures 2 and 4 of https://apps-nefsc.fisheries.noaa.gov/saw/sasi_files.php?year=2023&species_id=33&stock_id=6&review_type_id=5&info_type_id=5&map_type_id=&filename=2023_BSB_UNIT_WP_Truesdell_Curti_c_2023_Surveys.pdf suggest a higher CV. I could not find reports from the 2023 research track for black sea bass. The paper will be more relevant if CV=0.3 + OM_high is also investigated. Present results similar to how the low movement scenario was handled.
6. L194. A possible difference in SpD and SEP models is that SpD could have shared parameters across regions. Is this possible in WHAM, and if so was this used in the SpD Ems?
7. Equation 1. Explain what the truth is and how is it calculated. For example, in the feedback period, are the stock F's based on catch advice (no error) etc. Does the truth vary in each iteration of an OM, etc.
8. Check the accuracy of figure labels, and figure and table captions. I found a few errors but I did not examine the low movement figures closely.

Minor Comments

1. L39. I am sure how well NA REs account for these processes. That is the focus of this paper, wrt to movement. Maybe 'represent misspecification of' is better wording?

2. L50. I don’t think Cadigan (2016) claimed this. He just cited Gudmundsson and Gunnlaugsson (2012) who suggested that NAA REs could represent migrations. Maybe I misunderstood the text, and if so, this text should be clarified.
3. L59. Traditional assessments models don’t say anything about spatial distribution. They just model the stock as a whole. I think the same for demographic homogeneity. We know this exists but we just model population totals with average weights, maturities, etc. I suggest the authors refine this text to clarify their points.
4. L71-72. Seems like a contradiction. Please clarify what you mean.
5. L109. “in a controlled, research-based environment”. I don't see the value of this text? Isn't this always the case for MSE? Please clarify your point.
6. L172. clarify if the movement rates varied in each simulation iteration.
7. Figure S1 and Figure S16: Describe panels and clarify in the caption that the same movement rates are used in the spring and winter (I think).
8. L217. Describe why only recruitment random effects were assumed to continue in projections?
9. L233. I assume the WHAM models were fit 11 times for each iteration of an OM: 1st after 30 year historical period, then ten times in 30-year feedback period? It will help clarify the MSC if the authors should describe this.
10. Fig. S2 and S3 use different acronyms for EMs. Should use the ones in Table 3.
11. L304. I think this should be Fig. S2.
12. L305-307. Describe the basis for this conclusion.
13. Fig. S5. IQR is commonly defined as the difference between the 75th and 25th percentiles, and not the 40^{th} and 60^{th} percentiles like in this figure. Revise this.
14. Fig. 6 and S11 have the same caption?
15. L360. Not lower than the SpE Ems, correct? Clarify this.
16. Fig.7. Caption should clarify that the number in each panel is the total score.
17. Figures S14-S15. fix OM_low in caption.
18. L568. Mention that differences in maturation and growth rates are key biological characteristics.

Noel Cadigan