Center for Independent Experts (CIE) Peer Review of the Atlantis Ecosystem Model in Support of Ecosystem-Based Fishery Management in the Gulf of Mexico Large Marine Ecosystem

Review Panel Summary Report

April 2023

Introduction

A Review Workshop was convened during March 28-30, 2023, at the Florida Fish and Wildlife Research Institute in St. Petersburg, Florida to evaluate the performance characteristics and to identify appropriate management applications of an Atlantis ecosystem model employed by the University of South Florida (USF) to support the Southeast Fisheries Science Center's (SEFSC) evaluation of Ecosystem-Based Fishery Management (EBFM) strategies for the Gulf of Mexico (GOM) Large Marine Ecosystem. This review was undertaken as part of an EBFM-funded project at the SEFSC.

This report provides combined review comments—i.e., from both CIE and regional scientific reviewers—for each of the review objectives listed in the reviewer report template distributed as part of the review workshop briefing book. More detailed review comments addressing each of the review's terms of reference are being provided separately in the individual CIE review reports.

Meeting Facilitator
Dr. Matt Freeman (Gulf of Mexico Fishery Management Council).

Review Panel
CIE Reviewers: Dr. Vidette McGregor, Dr. Daniel Howell, and Dr. Ken Drinkwater.
Regional Reviewers: Dr. Luiz Barbieri, Dr. Joshua Kilborn, and Dr. Dave Chagaris.

Project Team
The project's Principal Investigators (PIs) and Co-PIs are Dr. Michelle Masi (SEFSC/SERO), Dr. Cameron Ainsworth (USF), Dr. Isaac Kaplan (NWFSC), Dr. Howard Townsend (OST), Dr. Skyler Sagarese (SEFSC), Dr. Chris Kelble (AOML), and Dr. Mandy Karnauskas (SEFSC).

The modeling team is composed of Dr. Cameron Ainsworth (USF), Dr. Holly Perryman (USF/IMR), and Rebecca Scott (USF).

Review Panel Comments

Objective 1 is to evaluate the data, parameterization, and skill of the GOM Atlantis model, with emphasis on predicting stock dynamics and catch of Penaeid shrimp (Brown, White and Pink Shrimp groups) and major interacting species.

Data

- Initial model conditions were set based on 2010 data. Model run was initially defined based on 1980 with some data gaps, and this was run forward to 2010 to tune and obtain parameter values.

- While the Review Panel was not provided with a thorough review of data inputs for all groups, numerous issues were discovered in the report and raised during the Review Workshop. Most of the issues stem from a lack of what the Review Panel referred to as "sanity checks", i.e., basic scrutinization of model inputs/outputs that should have been conducted prior to the CIE review. For example, Table 1 of the draft 2023 Tech Memo indicates that in GoM Atlantis, menhaden and large sharks have approximately the same biomass (2 million mt), snook and pinfish exist in almost equal biomass (194,294 mt), and there is more biomass of blacktip shark (54,096 mt) than macroalga (50,824 mt). Landing inputs, especially for the recreational fleet, were incorrectly calculated because they only include harvested fish (MRIP data types A plus B1) from MRIP in 2020. All B2 (i.e., discards) data from MRIP and the Texas, Louisiana, and Mississippi creel surveys were inadvertently excluded. Additional issues pertaining to Penaeid shrimp are discussed in more detail below.

- In general, oceanography is forced in an Atlantis model. For this particular model implementation simulations beyond the training period (1980-2010) are based on outputs from an oceanography model representing 2012. The Review Panel believes this generates a temporal disconnect in the hydrodynamic coupling and that the model needs a longer series of oceanographic forcing if it is going to be considered representative of more realistic environmental scenarios for the Gulf of Mexico..

- The data needs of any Atlantis model are extensive, and not all of the desired data will be available, and that is no different in this Gulf of Mexico implementation. There are some regional biases in the gut content data, as this has primarily been collected by Florida programs. It may be useful to incorporate more spatially explicit diets (e.g., diet data for the western vs. eastern Gulf of Mexico). Further, the data for seagrass habitats, one of the major focus areas for improvement to the model, appear to be somewhat outdated. More modern seagrass coverage metrics should be employed where possible.

Parameterization

- The model was in some respects mis-parameterized and for some species far from reality. The primary recommendation is that the modeling team consult with empiricists and regional stock assessment scientists to ensure the model is grounded in reality. During the Review Workshop it was stated that a series of review workshops were held with the SEFSC, but the nature of those reviews was not discussed. Given that those reviews did not catch obvious issues with the model, a different approach should be taken to ensure the model is scrutinized internally before going for further review. It is therefore not possible to recommend specific future research and data collection in support of the model; not until major issues have been resolved and we can distinguish between errors created when deriving model inputs and model uncertainty associated with ecological processes.

- Depth levels appear somewhat arbitrary and possibly not refined enough throughout the upper portion of the photic zone. Current maximum depths for each bin are: 10, 20, 50, 200, 2000, and 4000 m. The degree to which this is appropriate or not will depend not just on model dynamics but also on the specific question being posed.

- The diets were based on literature but could greatly benefit from receiving input and validation from local experts. Examples of key errors identified during the Review Workshop: red snapper eating snook, significant consumption of seagrass by spotted seatrout, and dinoflagellates overrepresented in diets of some species.

  Perhaps use of the Dirichlet method produced unrealistic diets. However, given the limited spatial coverage of the diet data and the number of problems identified during the Review Workshop the Review Panel feels that all of the diet data are to be considered for a full review with increased scrutiny. For example, the overarching approach of "never removing potential prey items" from the diet matrix is likely problematic, and some linkages simply should not be entertained (see itemized examples above). The Review Panel recommends that all diets (i.e., input and realized) be reviewed in consultation with local expert knowledge.

- Life history parameters also appear to need greater input from local experts. For example, shrimp (all three species) have been parameterized with incorrect spawning seasons (for both timing and duration of spawning). Furthermore, the vast majority of spawning windows in the model are probably too short, as the region has been shown to contain many species with protracted spawning periods and bimodal peaks annually. Thus, much like the diet data, the parameterization of spawning windows should be closely reexamined. It may be more informative (and realistic) to define these spawning periods using appropriate probability distribution functions across wider temporal windows (and with multiple peaks as necessary) to capture the dynamics in this highly unique environment.

- Parameterization of longevity for the three focal species of Penaeid shrimp has also been identified as a major concern (i.e., they need to live for c. one year, not 10 years). The Review Panel feels having this issue resolved is critical before this Atlantis model is able to produce realistic outputs.

- Parameterization of key species that interact with Penaeid shrimp should also be evaluated. The modeling team presented a list of 10 key species, apparently identified by the SEFSC during their reviews. The review panel lacks the information to comment on how appropriate this choice is, and in any event it should be an iterative process with the list revised as the links within the Atlantis model become clearer. Of the key species, model errors were identified that allowed red snapper to eat snook in large proportions, and seatrout and red drum were eating a large amount of seagrass. These issues likely stem from data processing errors, and the modeling team was recommended to revisit the source data to confirm the diet compositions. Also, the age structure, and the choice to have the oldest age group as a plus group or not, should be evaluated for all species..

- Fishery selectivities need to be reconsidered with further input from local experts. Knife-edge selectivity has been used for all selectivities and these might not be appropriate in all cases.

Skill

- Model skill was not evaluated for any species and the modeling team did not provide any fits to observed data. Rather a series of diagnostic tests and stability checks, based on equilibrium projections, were provided for shrimp and the 10 key species. The model failed several of the diagnostics. In the case of using MSY estimates from GoM Atlantis as a diagnostic, the rates derived from EwE models and used in comparison were incorrectly calculated, and the MSY estimate from Atlantis had not been compared against reported catches to "sanity check" the estimate. Additional information on individual body weights for the key species would have helped the review panel understand what was going on, as well as simple comparisons with observed data and known values (i.e., "sanity checks"). Diagnostics for other species were not provided or conducted, but they should at least be evaluated for basic stability and dynamic checks.

- Once the model is ready for skill assessment, examples of key areas that should be further assessed include: (1) realized diets against realistic diet inputs, and (2) realistic life history parameters (growth, longevity, reproduction, natural mortality) rates compared to best knowledge (based on literature and local experts if both available). Where other comparisons are possible (e.g., MSY estimates against catch data and assessment model estimates) these should also be conducted.

Objective 2 is to identify the extent to which the GOM Atlantis model is suitable for incorporating environmental effects relevant to shrimp production.

- The Review Panel feels that in its current form this Gulf of Mexico Atlantis model is not suitable for evaluating the role of environmental effects on Gulf of Mexico Penaeid shrimp production. For Penaeid shrimp, ecosystem drivers can be extremely important, ecosystem evaluations are definitely important, and the Atlantis framework does have potential in this space. However, a number of biological and ecological factors—particularly the data and parameterization issues mentioned above—for this particular model need to be addressed before any judgement on the ability of the Atlantis model to model these drivers. The focal Penaeid shrimp species need to be modeled with more realistic life history parameters and the model as a whole needs to be tested and run with more and varied environmental forcing than only one year (currently only 2012). Follow-on effects from environmental forcing will also need to be tested as these scenarios are developed.

Objective 3 is to determine the readiness of the model to conduct simulations that assess ecosystem-level impacts of climate change. This could include representation of habitat changes, changes in environmental conditions, and tolerances of species.

- Similar to comments provided above for Objective 2, the Review Panel feels that this GOM Atlantis model is not ready for simulations that assess ecosystem-level impacts of climate change. From a habitat perspective, seagrass has been included in the model with additional complexity in structure and growth rates. However, in this model implementation seagrass has been forced in the model, so it is not actually responding to changes in the system. This will need to be further explored before this

model can be considered for evaluating different environmental and climate change scenarios.

- Evaluation of changes in environmental conditions and species tolerances were essentially just efforts to determine if the modelers could get the model to respond to forced changes in the temperature and seagrass attributes. These "proof-o- concept" attempts did produce changes, however the changes that were produced did not seem realistic. In particular, large range shifts were noted for Pink Shrimp that did not seem likely (i.e., from extreme S. Florida to extreme N. Florida as the center of the species distribution). While it is encouraging that these initial experiments produced results, the code and functionality of the model need to be fully vetted, and each particular scenario examined will need to be checked to ensure that the model is likely to be producing realistic answers to any given specific question.

Objective 4 is to review recent updates to the Atlantis code base specific to the GOM Atlantis model which improves representation of seagrass dynamics. A novel routine was developed in 2021-2022 with CSIRO Australia. The routine partitions seagrass using pseudo age structure to improve representation of herbivory. The review will not otherwise focus on the Atlantis code base, nor will it focus on data quality except as it pertains to model performance.

- The novel seagrass routine was implemented to include age-structure on seagrass and allow for different modes of grazing (on leaves vs rhizomes). This is clearly an improvement over the previous modelling of this important habitat within the GOM Atlantis model. However, this novel routine was described, but not demonstrated. Only seagrass forcing scenarios were presented during the CIE review workshop. The Review Panel feels that the biggest threat to GoM seagrasses is water quality, and more effort should be put on improving representation of those processes. One challenge for GoM Atlantis is that the spatial resolution of the polygons is too coarse to capture fine scale patchiness and distribution of seagrasses.

Summary of Review Panel Comments and Recommendations

- The key recommendation is that more effort should be placed on comparing model results to available data and expert knowledge to identify where the model does and does not perform realistically. Specific examples are provided below, but the comparison should not be restricted to these, rather the researchers should be comparing the "reality" of the model whenever possible.
- The modeling team should provide a conceptual model for how Penaeid shrimp interact with their environment and fisheries, key drivers, and the spatial-temporal scales at which they take place. Then, demonstrate how those processes are represented in Atlantis.
- Fix shrimp age structure, growth, and spawning seasonality to better reflect what was described as their life history.
- Develop a way to post-process Atlantis output to obtain historical estimates of natural mortality and fishing mortality. This is one type of output that could be used as input to a stock assessment and should be trivial to produce from the Atlantis outputs.

- Conduct an ecological sensitivity analysis to identify key species linked to shrimp (or whichever is the focal species). For example, which groups in Atlantis respond the most to changes in shrimp biomass? Compare this to the pre-defined set of key species and refine that set if necessary.
- For long-lived fish species, consider increasing the number of age classes and/or making use of the plus-group option in Atlantis to allow for better representation of key ontogenetic and fishery selectivity processes, and better linkage with age-structured stock assessments. This should be especially considered where mortality due to organisms aging out of the age structure (effectively a poorly constrained senescence mortality) is an important driver of stock dynamics.
- Diet inputs were questionable, even for species with sufficient data. The modeling team should revisit the diet matrix, drilling down to raw data where there are inconsistencies with reality (based on basic knowledge of species overlaps and morphology).
- The modeling team should more closely evaluate the biomass inputs, especially across trophic levels and in relation to known biomass estimates from stock assessments. They should spend more time developing and evaluating PreBal diagnostics to identify and resolve errors. For example, in this implementation of GoM Atlantis large sharks and menhaden have equal biomass, there is an equal amount of pinfish and snook, and blacktip shark have higher biomass than macroalgae. These are obvious errors that should have been detected sooner.
- Only recreational landings data from MRIP type A+B1 catch in 2020 were included in the current GoM Atlantis model. This only represents known or reported harvests from the eastern Gulf—the majority of discarded fish were excluded. Data should be included from Texas and Louisiana, as well as the MRIP B2 along with discard mortality rates from stock assessments. These data were apparently provided by the SEFSC but excluded at the discretion of the modeling team. Also, exercise caution in using 2020 data for initial parameterization, and do not rely entirely on this year, as some of those data were imputed due to sampling pauses during covid.
- GoM Atlantis uses a combination of the larval dispersal source-sink matrix and default recruit-where-spawned assumption. This should be revisited on a species-by-species basis, as many species that spawn on the shelf distribute their larvae nearshore where they spend their early life history as juveniles. Also, the larval dispersion and transport, as well as any relationship with wind should be based on wind stress and not wind velocity.
- Examine why diet composition can change so drastically for some groups within the first year, as this will likely indicate spatial predator-prey mismatch issues. Use available knowledge of likely diets to identify which is the more realistic set of dynamics.
- Once a valid model is produced, validate emulator model runs with a few Atlantis runs based on the same inputs. This is important because uncertainty analysis is prohibitive in Atlantis due to long run time, so emulator runs could be a replacement, if they are demonstrated to be similar to actual model runs.