From guinea pigs to guides: advancing lessons learned from reviewing end-to-end marine ecosystem models

Holly A. Perryman[1*], Cameron H. Ainsworth[1], Michelle Masi[2], Isaac C. Kaplan[3], Howard Townsend[4], Skyler R. Sagarese[5], Matthew A. Nuttall[5], Rebecca L. Scott[1], and Hallie C. Repeta[1]

[1] College of Marine Science, University of South Florida, St. Petersburg, FL, USA
[2] Southeast Regional Office, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, St. Petersburg, FL, USA
[3] Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA
[4] National Marine Fisheries Service, Office of Science and Technology, Marine Ecosystems Division, Silver Spring, MD, USA
[5] Southeast Fisheries Science Center, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Beaufort, NC, USA

Abstract

Holistic approaches to managing living marine resources, which consider the suite of ecological and environmental interactions affecting a population, are becoming increasingly common. Often, these approaches utilize predictive models that include ecosystem dynamics. However, the application of marine ecosystem modeling tools is generally limited due to the lack of formal reviews of the model's utility and performance, despite these practices being commonplace for single-species stock assessment models. Herein, we provide an account of our experience undergoing a formal review of a Gulf of Mexico end-to-end marine ecosystem model. Guided by lessons learned from the "guinea pigs'' preceding us, described by Kaplan and Marshall (2016), we crafted and implemented a two-phase project timeline consisting of an informal review with regional experts and a formal review with independent experts. While the outcome of our review was that the model was not yet ready for use, a list of necessary model refinements provided by the reviewers offered a clear path for the model toward operational use. We reflect upon the practical challenges, successes, and setbacks encountered during our experience, offering insights into structuring a marine ecosystem model review for future applications. Additionally, building upon previous recommendations, we provide a list of baseline standards for reviewing marine ecosystem model performance. Addressing the inherent challenges in the review of marine ecosystem models is crucial for unlocking their potential contributions to ecosystem-based management, and our recommendations outlined herein offer guidance for future reviews.

Keywords

formal review; model performance; marine ecosystem models; Atlantis; ecosystem-based fisheries management

Code availability

Scripts and other materials pertinent to this article are freely available on GitHub: https://github.com/hollyannperryman/Gulf-Mexico-Atlantis-Peer-Review.git.

Correspondence:
Contact H. Perryman at holly.ann.perryman@dnv.com

Cite this article as:
Perryman, H.A., Ainsworth, C.H., Masi, M., Kaplan, I.C., Townsend, H., Sagarese, S.R., Nuttall, M.A., Scott, R.L., & Repeta, H.C.
From guinea pigs to guides: advancing lessons learned from reviewing end-to-end marine ecosystem models
Socio-Environmental Systems Modelling, vol. 7, 18848, 2025, doi:10.18174/sesmo.18848

Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

1. Introduction

There is growing appreciation that the management of living marine resources should shift from the traditional single-species approach to a more holistic approach, acknowledging that fishing is merely one element in the collection of natural and anthropogenic drivers that influence a population. Such an approach is often called Ecosystem-Based Fisheries Management (EBFM) (Brodziak and Link, 2002; Pikitch et al., 2004; Fogarty, 2014). To support EBFM, the predictive modeling tools utilized for research and advice must account for ecosystem interactions. This includes the development of more holistic fisheries stock assessment models that capture ecosystem dynamics (Lynch et al., 2018). In the USA, there has been notable progress in expanding single-species stock assessments to include ecosystem considerations (Marshall et al., 2019). Another avenue toward EBFM is the incorporation of information from more complex marine ecosystem models (MEMs). As delineated by Plagányi (2007), these include — but are not limited to — dynamic multispecies models (e.g., Gadget), dynamic system models (e.g., Ecopath with Ecosim), and spatially explicit dynamic system models (e.g., Atlantis). A notable success story from the Gulf of America (previously referred to as the Gulf of Mexico, hereafter abbreviated as GOM) is the inclusion of red tide mortality effects on population dynamics of economically important grouper species in the Southeast Data, Assessment, and Review (SEDAR) stock assessment process (SEDAR, 2019; SEDAR, 2021; Sagarese et al., 2021). In this case, an age-specific index of red tide mortality for gag grouper (*Mycteroperca microlepis*) was derived from the West Florida Shelf Ecopath-with-Ecosim (EwE) and Ecospace model (Vilas et al., 2023) and used to inform the potential effect of the 2021 red tide event on gag grouper during catch projections (SEDAR, 2021). This application is the first case in the GOM of a MEM contributing directly to the development of catch advice, giving inspiration for applications to other stocks and with other MEMs.

A key factor limiting the application of MEMs for the development of fisheries management advice is the uncertainty regarding model utility and credibility (Link et al., 2010) in conjunction with the lack of model performance evaluations (Steenbeek et al., 2021; Craig and Link, 2023; Kempf et al., 2023), which also hinders the adherence to good modeling practices (Jakeman et al., 2024). Stock assessments undergo a formal review process to communicate technical details and evaluate methodology, ensuring that model results are providing the best available insights for effective management (Lynch et al., 2018). We defined a formal review as a structured evaluation process in which one or more qualified experts assess the scientific information to ensure its quality and credibility. The experts, who must be independent from the model's development and any affiliations with agencies or constituent groups, must conduct an impartial, objective review, free from conflicts of interest and external influence. Formal reviews in this context, and in our discussion below, typically involve panels of three or more experts who spend multiple days learning about model dynamics and questioning model developers; in this manner they exceed the scope of review, verification and assessment required for scientific journal publications (though such scientific journal reviews often precede the formal review panel).

Formal reviews of MEMs are infrequent (Steenbeek et al., 2021; Craig and Link, 2023), though some cases have been documented. For example, the stochastic multi-species modeling methodology was reviewed by the International Council for the Exploration of the Sea (ICES) (ICES 2016, 2019c, 2023). An EwE model of the Irish Sea was reviewed informally (ICES 2018, 2019b) and formally (ICES, 2019c). Additionally, a formal review of multiple models of mixed complexity was attempted during the Atlantic menhaden stock assessment (SEDAR 2020). Lastly, an Atlantis model of the California Current was formally reviewed through the National Oceanic and Atmospheric Administration's (NOAA) Center of Independent Experts (CIE) (Kaplan and Marshall, 2016). Considering the need to formally review MEMs and the current rarity of formal reviews of MEMs, it is imperative to leverage these cases for progress.

There is a diverse toolbox of MEMs available for the GOM (O'Farrell et al., 2017), including a biogeochemical end-to-end MEM of the Gulf of Mexico (GOM Atlantis). The GOM Atlantis model was developed to support EBFM (Ainsworth et al., 2015; Perryman et al., 2023b) and has been used to investigate the implications of oil pollution (Dornberger et al., 2023; Morzaria-Luna et al., 2022; Ainsworth et al., 2018), point-source nutrient introduction (Dornberger et al., 2023), ecosystem indicators (Masi et al., 2017), management strategy evaluation (Masi et al., 2018), and climate change (Olsen et al., 2018). The Atlantis modeling framework is spatially explicit and built on dynamically integrated submodules for physical and biogeochemical processes, ecology, human uses, and management (Audzijonyte et al., 2019). Atlantis is a complex MEM, and its complexity can reach even higher levels when optional features are activated (e.g., environmentally driven ecological responses, dynamic assessments). While increased complexity may improve ecological realism, it comes at the cost of increased model uncertainty (Collie et al., 2016).

Our project executed a formal review of the GOM Atlantis model with the goal to assess the model's readiness for providing EBFM advice for GOM shrimp (penaeid) stocks. We conducted a formal review in a manner similar to a stock assessment review (Lynch et al., 2018), and leveraged experience from a previous Atlantis model review (Kaplan and Marshall, 2016). This was an extensive process for both modelers and reviewers which culminated with a list of necessary model refinements provided by the reviewers, offering a clear path for the GOM Atlantis model toward operational use in support of shrimp EBFM. Leveraging the experience described in Kaplan and Marshall (2016) was invaluable, underscoring the importance of sharing lessons learned to help advance the broader community by developing frameworks that assist both modelers in preparing for reviews and reviewers in executing them. In this paper, we overview our experiences and build upon the lessons learned by Kaplan and Marshall (2016), marking the transition from being test subjects (i.e., “guinea pigs”) towards establishing guides to support and streamline future MEM reviews. First, we overview the GOM Atlantis model review. Next, we summarize lessons learned, including: structuring a MEM formal review, standards for evaluating MEM performance, and accounting for uncertainty. Lastly, we provide supplementary details of our MEM review process to further contextualize our lessons learned.

2. How we got here: an overview of the Gulf of Mexico Atlantis model review

Leveraging lessons learned by Kaplan and Marshall (2016), we focused the review of the GOM Atlantis model on a specific application (providing EBFM advice for Gulf shrimp) to give the review clear context. To further direct the scope of the review, we concentrated on a general research topic (shrimp productivity scenarios) and limited diagnostic evaluations to shrimp groups and their major interacting species (key predators and food sources). Additionally, we developed and implemented a two-phase approach to execute the review.

In the first phase, we subjected the existing GOM Atlantis model to an informal review with regional experts at the National Oceanic and Atmospheric Administration’s (NOAA) National Marine Fisheries Service (NMFS). The informal review was held over a series of six webinars, with approximately one webinar occurring every one to three months (spanning a total of about ten months). The objective of the informal review was to prepare for the formal review, specifically: (i) identify major species interacting with Gulf-shrimp species, (ii) discuss and develop model diagnostics, and (iii) evaluate and improve the model's realism. For the second phase, we conducted a formal review of the GOM Atlantis model with a panel of independent experts from NOAAs CIE program (Brown et al., 2006), which we opted to supplement with three external (non-NOAA) regional experts. The formal review was held over a three-day meeting which was open to the public (March 28-30, 2023). The objectives of the formal review were the following: (i) to evaluate the data, parameterization, and skill of the GOM Atlantis model, with emphasis on predicting stock dynamics and catch of penaeid shrimp groups (brown *Farfantepenaeus aztecus*; white, *Litopenaeus setiferus*; and pink, *Farfantepenaeus duorarum*) and their major interacting species, (ii) to identify the extent to which the GOM Atlantis model is suitable for incorporating environmental effects relevant to shrimp production, (iii) to determine the readiness of the model to conduct simulations that assess ecosystem-level impacts of climate change (e.g. representation of habitat changes, changes in environmental conditions, and tolerances of species), and (iv) to review recent updates to the Atlantis code base specific to the GOM Atlantis model which improves representation of seagrass dynamics.

Ultimately, the CIE review panel concluded that GOM Atlantis is currently not ready for operational use in shrimp management but noted the model's potential to address such strategic questions. The reviewers provided a list of necessary model refinements, offering a clear path for the GOM Atlantis model toward operational use in support of shrimp EBFM. All documentation pertinent to the setup, execution, and results of the GOM Atlantis model review has been made available through a GitHub repository (Perryman, 2024). Our experience conducting the review shaped a set of lessons learned, which we share in the following section to support future MEM reviews.

3. Looking ahead: lessons learned from formally reviewing a marine ecosystem model

This section presents key lessons learned from our experience reviewing the GOM Atlantis model. These insights are intended to support future efforts in reviewing MEMs, particularly in a formal review setting. We focus on three main areas: structuring the MEM formal review, standards for evaluating MEM performance, and addressing uncertainty.

3.1 Structuring a marine ecosystem model formal review

Targeted focus

Concentrating the MEM review on a focal species and its key interacting groups allowed the review to dive deeper into the dynamics most critical to the project, rather than executing a surface-level examination across all the biological groups and fisheries represented in the MEM. Some MEMs simulate an extensive list of components. For example, GOM Atlantis includes 91 functional groups and 23 fishing fleets (Ainsworth et al., 2015). For these MEMs, executing an extensive performance evaluation across all modeled components in a single review would be an unreasonable task for both modelers and reviewers. Through this application, we also learned that selecting focal model components is not straightforward. Collaborating with regional scientists during the informal review was valuable for selecting focal groups; however, regional scientists in the formal review later questioned these choices. This underscores the need to provide a strong justification for these selections. Additionally, preparing supplementary materials to facilitate the review of other groups and model components, should they be requested during the formal review, would be prudent.

Two-phase approach

Implementing an informal review with regional experts demonstrated considerable value in the preparation for formally reviewing a MEM. Other MEM reviews have executed similar two-phase approaches (e.g., ICES, 2018, 2019b,c). Significant advantages of this approach include assistance in developing visuals to convey model diagnostics, which is important for the acceptance of complex modeling results (Rodriguez-Perez et al., 2023), and the efficient identification and prioritization of model refinements essential for achieving project milestones. Although our two-phase approach had its advantages, we have identified several revisions to the informal review process that could benefit future MEM reviews.

First, model documentation was submitted for examination prior to the formal review (see Section 3), but this was not done for the informal review. Offering this documentation at the start of the informal review would have been helpful to the regional experts involved. Second, while the informal structure allowed for dynamic discussions, adding more structure through clear objectives from the outset could result in greater outputs. For instance, a formal review with regional experts, particularly managers, could help the modeling team identify the most relevant applications and uncertainties for the formal review. Lastly, the experts participating in the informal review found it challenging to keep up with model changes over the extended timeline. A shortened informal review timeline could benefit participants; however, interactions with a regional team of experts should continue beyond the informal review process, putting the recommendation for frequent engagement (Townsend et al., 2019; Craig and Link, 2023) into practice. This underscores the importance of carefully considering the review timeline (discussed further below).

Regional Expertise

During the formal review, the insights regarding local ecosystem dynamics provided by three supplemented regional experts proved invaluable to the CIE reviewers, who were less familiar with the GOM ecosystem. With respect to the evaluation of the model, our experience suggests that regional reviewers are well-suited to evaluate region-specific parameters and data (e.g., group spatial/temporal trends, diets, fleet dynamics, environmental patterns). Conversely, CIE reviewers are well-suited to evaluate general dynamics and model assumptions (e.g., model stability, natural mortality, productivity). This division of labor between regional and CIE reviewers could improve the evaluation process, ensuring comprehensive scrutiny of both localized nuances and broader ecosystem dynamics. Additionally, the formal review was enriched by the insightful regional feedback provided by public attendees. Future MEM reviews should ensure public participation (e.g., through virtual access and engagement; Abas et al., 2023), not only to tap into regional expertise but also to promote communication and instill confidence among stakeholders.

Careful consideration of timeline

Based on our experiences, the formal review process for a MEM would benefit from following a more interactive and longer timeline than that of fisheries stock assessments. In the GOM, stock assessment reviews are managed under the SEDAR process, typically involving a week-long workshop with a panel of independent expert reviewers. A week-long in-person workshop has proven adequate for reviewing data inputs, model configuration, and uncertainty analyses of a stock assessment, however attempting a formal review of the GOM Atlantis model in a similar manner proved difficult, despite focusing the scope of the review to components pertinent to penaeid shrimp rather than the entirety of GOM Atlantis dynamics. Challenges included having ample time to properly introduce reviewers to the pertinent dynamics, for reviewers to thoroughly examine and discuss the many diagnostics, and to produce additional model output to follow up on any requests from the reviewers. Similar challenges were noted following a recent Atlantic menhaden (single species) stock assessment which also considered ecological reference points (Howell et al., 2021; Reum et al., 2021).

Under SEDAR, the assessment process for new analytical approaches ('benchmark' assessments) is organized around a series of workshops with both regional and independent experts, focusing on analyzing and reviewing datasets, developing and refining quantitative population analysis, and conducting an overall review. Extending the MEM formal review timeline to something similar could be challenging for independent experts who often must travel to attend in-person meetings. However, the CIE review program consists of three types: desk (a brief virtual call for reviewers to ask questions), virtual panel (a full panel review held virtually), and in-person panel (a full panel review held in person). While fully virtual reviews are possible for stock assessments, they can reduce the time for reviewers to fully grasp the scope of the review (Lynch et al., 2018). Given the complexity of MEMs and the varying expertise levels of reviewers and attendees, a fully virtual formal review modeled after a SEDAR benchmark assessment is unlikely to be feasible for a MEM. Therefore, rather than being exclusively in-person or fully virtual, the formal MEM review process could effectively combine both formats.

For example, prior to the in-person workshop, an introductory virtual webinar could be held. This time could serve as to familiarize the reviewers with the scope of the review, the MEM, and introduce the submitted documentation. After a brief period of time (e.g., a couple of weeks), the in-person workshop could be held. Since introductory presentations have recently occurred, the in-person workshop could immediately commence the in-depth review of the model diagnostics and uncertainty, along with starting an inventory of any panel requests for additional analyses. Following the in-person workshop, a final virtual webinar could be held to serve as a follow-up on any panel requests and any final deliberation. We acknowledge that expanding the review process of a MEM may introduce additional logistical obstacles, such as increased time, effort, and expenses. Additionally, an extended time commitment may reduce the available pool of qualified reviewers. Similar to stock assessments, these challenges warrant consideration to ensure quality and credibility in the tools being used to develop fisheries management decisions (Lynch et al., 2018). Careful thought is necessary to develop a MEM formal review structure and timeline that strikes a balance between feasibility and productivity.

Concise summaries of model information

Formal reviewers are obligated to examine all documents submitted for the review, which can be numerous and lengthy. Our experiences illuminated the value of a succinct presentation of materials. Organizing information on the model's structure, development, parameterization and behavior into easily digestible summaries facilitates the comprehensive review process. This, in turn, improves the prospects for the model to be accepted in management applications. For example, throughout the review of GOM Atlantis, particular attention was paid to the fisheries subroutine and harvest, which offers a broad range of features (Audzijonyte et al., 2017b, 2019). Ultimately, the review resulted in updated parameterizations (following the informal review) and recommended improvements to fleet dynamics (following the formal review). These discussions, however, as well as discussions of other modeled subroutines, would have been better facilitated and streamlined with some sort of subroutine summary.

Summary protocols should be extended beyond model output diagnostics to also review individual components (e.g. species-level input data as well as output diagnostics) of the model, especially those directly pertinent in the review. For example, in the southeast region, "one-pagers" summarizing the utilized data and analysis results have become popular amongst stakeholders, as they succinctly summarize the key aspects of an analysis without requiring one to read an entire technical document. For a MEM review, "one-pagers" (or multiple-pages depending upon the content) could be prepared for each functional group relevant to the question at hand to summarize model inputs and outputs pertinent to the evaluation standards (see Section 3.2), allowing reviewers and interested stakeholders to easily and quickly identify potential issues in need of addressing. For example, 10 one-pagers could have been developed for this application covering Gulf shrimp, their major predators, and their prey.

Although not explicitly followed in this application, recent years have seen the development of protocols for summarizing MEMs that are aimed at enhancing transparency and reproducibility while providing a concise model overview. Examples include three screening questions proposed by Grimm et al. (2020b), the ODD protocol (Overview, Design concepts, Details; Grimm et al., 2006, 2010, 2020a), and the OPE protocol (Objectives, Patterns, Evaluation; Planque et al., 2022). Grimm et al. (2020b) proposed three critical questions for evaluating MEMs: the model's purpose, organization, and evidence of functionality. While these questions offer simplicity and effectiveness, the ODD protocol provides additional details on model design. Originally developed for agent-based models, the ODD protocol is increasingly used in the MEM community, leading to the development of the complementary OPE protocol by Planque et al. (2022), which includes evaluation procedures. Future model reviews should more formally and directly consider these protocols, even though some recommendations are similar to those items detailed in Kaplan and Marshall (2016).

3.2 Standards for evaluating marine ecosystem model performance

Evaluating MEM performance is necessary to foster confidence in its operational use (Steenbeek et al., 2021; Craig and Link, 2023; Kempf et al., 2023). This process is often referred to as model skill assessment or validation, which evaluates how well a model reflects the true system (Olsen et al., 2016). However, because true system dynamics cannot be directly measured, model performance evaluation is instead based on how well outputs fit to observational data (Stow et al., 2009). While there is uncertainty in both observational data and MEMs, together they offer a better depiction of reality (Skogen et al., 2021).

MEM performance is evaluated using metrics and characteristic signatures, for which there is a growing library. Stow et al. (2009) reviewed metrics and approaches for evaluating model skill assessment, specifically regarding ecological or biogeochemical MEMs coupled to a physical model. Link (2010) proposed a set of rigorous and uniform standards by which the foundation of all ecological models and related applications for any system can be assessed (termed the PREBAL diagnostics). Bennet et al. (2013) reviewed numerical, graphical and qualitative methods for assessing model performance techniques across various fields of environmental modeling. Kaplan and Marshall (2016) proposed credibility and quality control standards for end-to-end models. Hipsey et al. (2020) reviewed methods for evaluating the performance of aquatic ecosystem models. The ICES working group on multispecies assessment methods (WGSAM) proposed standardized criteria for consistent review of models (ICES, 2019c), which inspired a framework with guiding questions designed for specific, practical, and adaptable skill assessments of models for EBFM (Kempf et al., 2023). How one evaluates MEM performance, including the choice of metrics and methods of evaluation, depends on the MEM itself, the application scope, and observational data/information available (Stow et al., 2009; Bennett et al., 2013; Planque et al., 2022). For example, focusing model development and the review on Gulf shrimp and relevant species during review allowed a deep dive into model parameters, which will help guide future model refinements. A baseline set of standards for reviewing MEM performance improves efficiency, reduces ambiguity, and builds a foundation for interactive improvement.

In the following, we outline a set of baseline standards to consider when reviewing MEM performance. This list is based on our experience executing the recommendations from Kaplan and Marshall (2016), developing diagnostics through the informal review, and evaluating diagnostics through the formal review, as well as the consideration of the growing literature. There is no one metric that comprehensively captures all discrepancies between model outputs and observations (Stow et al., 2009; Olsen et al., 2016). Not only should a set of metrics be considered when evaluating MEM performance, but each modeled component should have its own collection of metrics. Our recommended set of baseline standards encompass evaluation criteria for (i) individual biological groups (Table 1), (ii) community structure (Table 2), (iii) modeled fisheries (Table 3), (iv) environment and group responses (Table 4), and (v) spatially explicit models (Table 5). We broadly outline our recommended set of baseline standards in the text, and provide tables that summarize examples of quantitative assessment methods, visual diagnostics, and reflections from our application (if applicable). Our recommendations serve as a base, and we advocate for customization to tailor to individual models and applications (Hipsey et al., 2020; Planque et al., 2022). Additionally, standards may need evaluation at various temporal and spatial scales. Any deviations from a standard should be explained to distinguish model weaknesses from exceptions (Link, 2010; Kaplan and Marshall, 2016).

i) Standards for the assessment of individual biological groups

While it is ideal to achieve these standards for all simulated biological groups, this can be challenging for end-to-end MEMs (Kaplan and Marshal, 2016; Pethybridge et al., 2019). A MEM review should meet these standards for biological groups pertinent to the model application The acceptance of any deviations should be left to the discretion of the review panel.

a. *Biomass trends persist and are relatively stable*. This standard is two-fold. First, no functional group should go extinct. Second, a simulation with no fishing and constant environmental forcing (no stochasticity) should result in no significant trends in forecasted biomass for the majority of vertebrate species/groups. These are general rules to demonstrate model stability, but any exceptions should be described (Kaplan and Marshall, 2016).
b. *Biomass trends display temporal patterns that qualitatively match data/expectation*. This standard is two-fold. First, modeled biomass trajectories for a historical period should reasonably align with available data. Second, modeled biomass trajectories should reproduce patterns of temporal variability, including magnitude and timing (e.g., migration, recruitment). Demonstrating that a model properly reflects temporal trends in biomass is essential to establish reasonable population dynamics (Kaplan and Marshall, 2016).
c. *Length/Weight at age qualitatively matches data/expectation*. Demonstrating that a model properly allocates resources to growth and appropriately scales it with size/age is essential to establish reasonable population dynamics (Hipsey et al., 2020; Kaplan and Marshall, 2016).
d. *Natural mortality decreases with size/age and is consistent with expectations*. Demonstrating that a model properly handles natural mortality and appropriately scales it with size/age is essential to establish a reasonable representation of life history and productivity (Kaplan and Marshall, 2016).
e. *Productivity reference points qualitatively match expectations*. Demonstrating that a model properly reflects the equilibrium relationships between catch and biomass is essential to establish a reasonable representation of stock productivity (Ainsworth et al., 2015; Kaplan and Marshall, 2016).

Table 1: Recommended baseline standards for evaluating the performance of individual biological groups, including assessment methods, reflections from our experience reviewing a marine ecosystem model (MEM), visual diagnostic examples, and literature examples other than Perryman et al. (2023b).

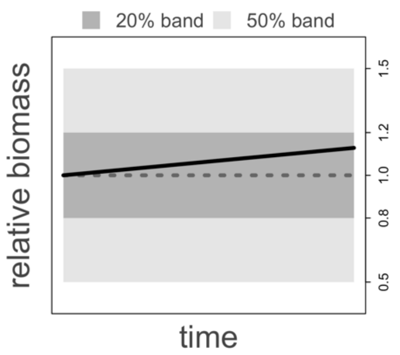
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Biomass trends persist and are relatively stable			
Persistence can be achieved when biomass is greater than 1% of initial biomass (Kaplan and Marshall, 2016).	The formal review considered a table indicating pass/fail, enabling a quick, foundational assessment across all groups.		Ortega-Cisneros et al. (2017).
Relative stability can be assessed visually with time series, and quantitatively with thresholds. In the case of Atlantis, the threshold range ±20-50% of initial conditions is commonly used (Horne et al., 2010; Audzijonyte et al., 2017a; Pethybridge et al., 2019).	The model should be simulated to equilibrium. For age-structured groups, cohorts should be examined individually. In Atlantis, vertebrate groups utilize nitrogen as the "currency", so evaluation should include biomass, abundance, and weight to avoid concealing underlying trends (Pethybridge et al., 2019). This was observed during the formal review of GOM-Atlantis, where apparent stability in biomass for red snapper (*Lutjanus campechanus*) masked opposing trends in abundance and weight (Perryman et al., 2023b).		McGregor et al. (2019); Pethybridge et al. (2019).

Table 1 (*continued*)

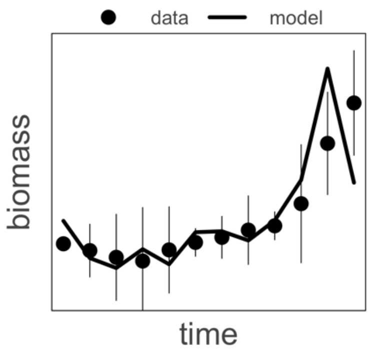
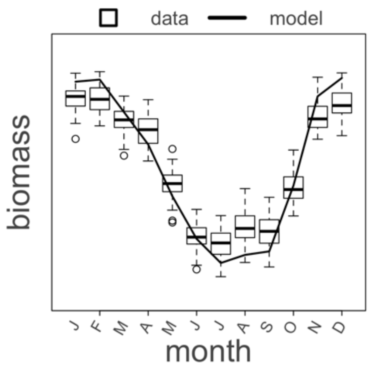
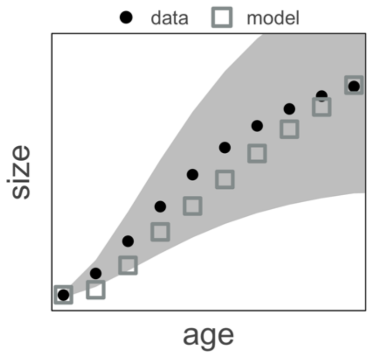
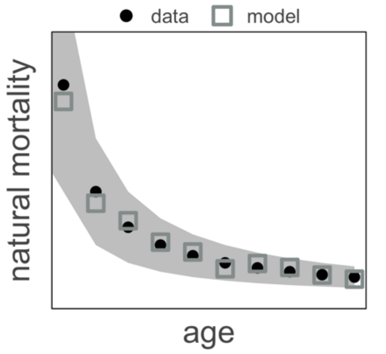
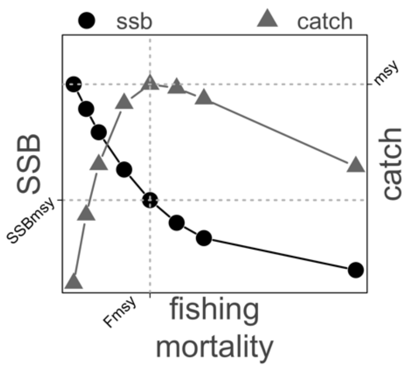
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Biomass trends display temporal patterns that qualitatively match data/expectation			
Visual inspection of time series, quantitative inspection of a combination of error metrics, and/or quantitative inspection of a correlation coefficient (Hipsey et al., 2020).	Ainsworth et al. (2015) presented visual comparisons of model outputs against observational data for the historical GOM Atlantis model. The formal review was of the "2016+" GOM-Atlantis model, a projection model which lacks a historical simulation.		Olsen et al. (2016); McGregor et al. (2019).
	The formal review considered visuals displaying annual variability, which allowed the inspection of the spawning seasonality of shrimps, and tables summarizing movement parameterizations, which allowed the inspection of migration dynamics.		
Length/Weight at age qualitatively matches data/expectation			
Visual inspection of the length-weight relationship, length-weight at age and/or size distribution histogram (Hipsey et al., 2020). Ideally, this should be achieved for the majority of vertebrate groups (Kaplan and Marshall, 2016). In the case of Atlantis, the threshold range ±50% of initialization has been used (Horne et al., 2010).	The formal review considered visuals depicting decadal model outputs against a region of acceptance (± 50% of initialization). These images sparked considerable discussion, underscoring the importance. Although the region of acceptance offered a quasi-skill assessment, reviewers encouraged future evaluations to prioritize comparisons against data.		McGregor et al. (2019).

Table 1 (*continued*)

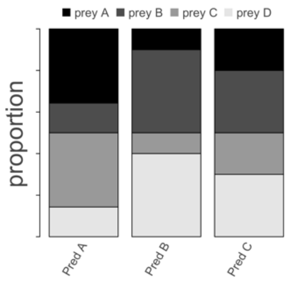
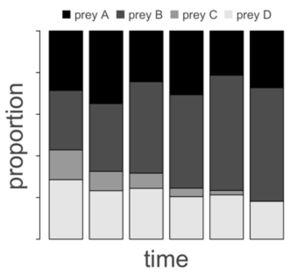
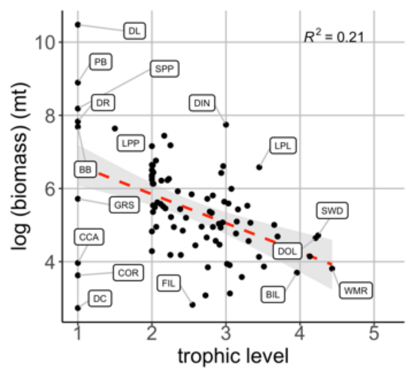
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Natural mortality decreases with size/age and is consistent with expectations			
Visual inspection, ideally by plotting natural mortality (including predation) as a function of age (Kaplan and Marshall, 2016; Hipsey et al., 2020).	In Atlantis, natural mortality is a combination of components (e.g., predation, density dependence, environmental impacts; Audzijonyte et al., 2019). Currently, Atlantis does not quantitatively report natural mortality for groups (Audzijonyte et al., 2017a). To evaluate this standard, the formal review considered visuals comparing abundance-at-age over time against the criteria that younger groups should have higher abundance than older groups. These images sparked considerable discussion, underscoring the importance. Although this approach allowed the evaluation of natural mortality dynamics, ultimately, reviewers recommended developing a method to quantitatively ascertain natural mortality from Atlantis. These metrics are not only important for model review but also constitute a crucial input for stock assessment.		McGregor et al. (2019).
Productivity reference points qualitatively match expectations			
Visual inspection comparing equilibrium relationships resulting from the MEMs against available stock assessments. For most groups (or 80% of biomass summed across all vertebrate groups) the productivity should qualitatively match productivity from stock assessments or life history theory (Ainsworth and Walters, 2015; Kaplan and Marshall, 2016).	Some MEMs have tools to predict equilibrium relationships (e.g., Ecosim allows the prediction of equilibrium yield for individual groups; Walters et al., 2005). For models lacking such tools, productivity can be evaluated by simulating the equilibrium response of biomass across a range of fishing rates. The assessment processes for Gulf penaeid stocks are currently being re-evaluated during a research track assessment, which is reconsidering model structure and how to best estimate reference points for this annual crop. Because of this, direct comparison of stock assessment estimates to Atlantis outputs were not feasible. Thus, during the formal review, Atlantis outputs were compared against outputs from a selection of EwE models.		Ainsworth et al. (2015); Rovellini et al. (2024).

ii) Standards for the assessment of community structure

a. Trophic interactions and realized diets qualitatively match expectations. Demonstrating that a model properly represents predator-prey dynamics is essential to establish a reasonable representation of system ecology (Kaplan and Marshall, 2016). This is exceptionally important given the significant influence that dietary parameterization has on model dynamics (Perryman et al., 2020).

b. Community structure qualitatively matches expectations. Demonstrating that a model properly reflects community dynamics (food web organization, trophic structure, biodiversity) is essential to establish a reasonable representation of ecosystem function (Hipsey et al., 2020). Additionally, properties pertaining to ecosystem community structure should be properly captured as MEMs can synergize with ecosystem indicators to support Ecosystem-Based Management (EBM) (Tam et al., 2019).

Table 2: Recommended baseline standards for evaluating the performance of the community structure, including assessment methods, reflections from our experience reviewing a marine ecosystem model (MEM), visual diagnostic examples, and literature examples.

Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Trophic interactions and realized diets qualitatively match expectations			
Visual inspection of plots or time-series or plots that show variance (Kaplan and Marshall, 2016; Hipsey et al., 2020). If the observational data are available, one should consider predator compositions for prey in addition to prey compositions for predators. Temporal diet dynamics should be considered as diets realized in-model may drift throughout a simulation.	The formal review considered visuals displaying the average species-specific prey compositions from simulation year 1 (quasi-parametrization) against average compositions throughout years 40-50 (quasi-equilibrium). These images distinguished juvenile and adult diets. These images sparked considerable discussion, underscoring the importance. We received a number of recommendations, such as considering spatially explicit predator-prey dynamics as to identify possible regional biases in either the observational data and/or model data. For predators with complex dietary linkages, although species-specific evaluations allow for detailed assessments, guild-specific evaluations may be acceptable and more easily digestible (e.g., Fulton et al., 2007).	 	Fulton et al. (2007). Porobic Garate (2019).
Community structure qualitatively matches expectations			
Hipsey et al. (2020) highlighted properties that assess modeled community structure (e.g., PREBAL diagnostics; Link 2010). Compilations of ecological indicators also offer methods for assessing ecological organization (Olsen et al., 2016; Coll and Steenbeek, 2017; Hipsey et al., 2020).	The formal review, proof-of-concept PREBAL diagnostics were shown as tools for developing PREBAL diagnostics with Atlantis outputs are still under development (Perryman et al., 2023b). The reviews encouraged the completion of such tools.		McGregor et al. (2019).

iii) Standards for the assessment of modeled fisheries

Evaluations of MEMs often focus on metrics for biological groups and community dynamics (Tables 1-2). While these criteria are undoubtedly crucial for evaluating a model, it's imperative to not overlook the evaluation of simulated fisheries. This aspect is particularly significant as it demonstrates the model's utility as a tool for testing fisheries management scenarios and providing advice for fisheries management. If applicable, fisheries outputs should be partitioned by categories pertinent to the system/application (e.g., landings, discards).

a. Catches qualitatively match data/expectation. This standard is two-fold. First, modeled catch trajectories for a historical period should reasonably align with available data. Second, forecasted catches should qualitatively agree with expectations. Demonstrating that a model properly represents catches is essential to establish a reasonable representation of ecosystem dynamics and function (Hipsey et al., 2020).

b. Catch structure qualitatively matches expectations. Demonstrating that a model properly represents catch structure (catch-at-age, fleet catch compositions) is essential to establish a reasonable representation of ecosystem dynamics and function (Hipsey et al., 2020).

Table 3: Recommended baseline standards for evaluating the performance of modeled fisheries, including assessment methods, reflections from our experience reviewing a marine ecosystem model (MEM), visual diagnostic example, and literature examples.

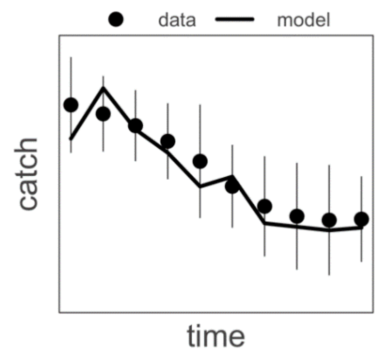
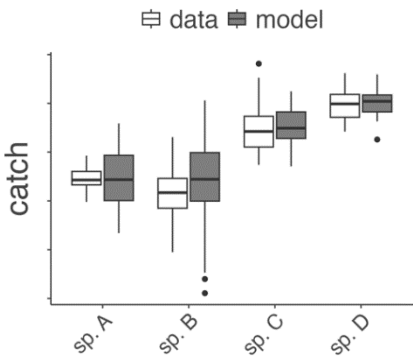
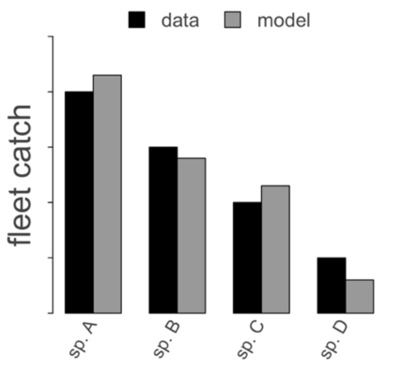
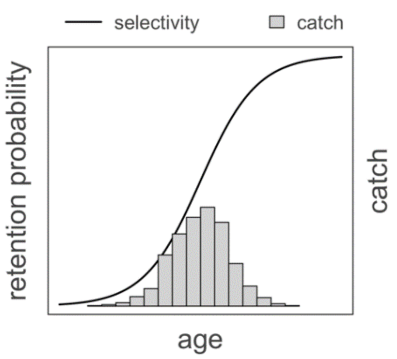
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Catches qualitatively match data/expectation			
Visual inspection of time series or the computation of a user-determined combination of error metrics (Hipsey et al., 2020).	The formal review did not cover historical model fitting, as the current application was focused on the second stage of model development - honing the present-day model for shrimp perturbation scenarios. Note, there is value in considering catches both in relative terms (rate) and absolute terms (tons). Considering rates is useful in terms of evaluating stock productivity, while the model's ability to match catch tonnage lends confidence that the scale of simulated stock sizes is adequate to support the magnitude of catches.		Olsen et al. (2016).
	The formal review considered a table that compared modeled average catches against the comparable recent observation data. This table distinguished commercial landings, commercial discards, recreational landings, recreational discards, and fishing mortality (Perryman et al., 2023b). While this condensed a lot of information into one medium, visuals would facilitate the evaluation process.		

Table 3 (*continued*)

Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Catch structure qualitatively matches expectations			
Catch structure refers to how catches are arranged and relate between parts of the system. Visual inspection can consist of time-series or plots that show variance. Additionally, there are a variety of catch indicators that can be validated against observation data (Olsen et al., 2016; Coll and Steenbeek, 2017; Hipsey et al., 2020).	The informal review considered fleet-specific catch compositions, which resulted in some model refinements. Visuals were provided to the formal review, but there was not enough time to delve into a rigorous formal review during the workshop. For mixed fisheries with diverse catches, evaluations may benefit from structuring the visuals so that key species are identified while the remaining organisms are classified by guilds. One should consider the distribution of a species' catch amongst the simulated fisheries in addition to fleet-specific catch compositions.		
	Formal reviewers recommended the consideration of presenting fleet selectivities and catch-at-age distributions, considering the importance of catch-at-age for some stock assessment methods.		

iv) Standards for environment and group responses

There is growing interest in using MEMs to gain insight and advice regarding the potential impacts of changing environments on ecosystems and management (Bryndum-Buchholz et al., 2020; Tittensor et al., 2021; Morell et al., 2023; Perryman et al., 2023a). A model review of such an application should include the evaluation of the handling of the environment, environmental changes, and subsequent responses from the biological groups. Below are some baseline standards based on the review from Hipsey et al. (2020).

a. The handling of physical components should qualitatively match expectations. Demonstrating that a model properly represents key physical components (e.g., temperature) is essential to establish a reasonable representation of the environment.
b. The handling of water quality and biogeochemistry should qualitatively match expectations. Demonstrating that a model properly represents key biogeochemical components (oxygen, nutrients, organic matter, pollutants) is essential to establish a reasonable representation of the environment, nutrient cycling and/or sediment processes.

c. The handling of group responses to the environment should qualitatively match expectations. Demonstrating how groups respond to environmental changes is essential to establish the model as a reasonable tool for testing scenarios of changing ocean conditions.

Table 4: Recommended baseline standards for evaluating the performance of the environment and group responses, including assessment methods, reflections from our experience reviewing a marine ecosystem model (MEM), visual diagnostic example, and literature examples.

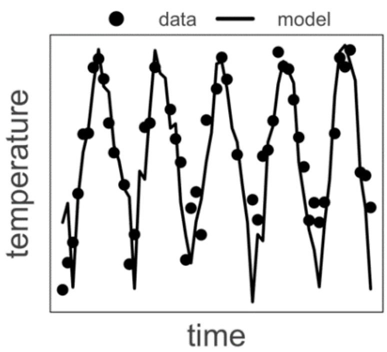
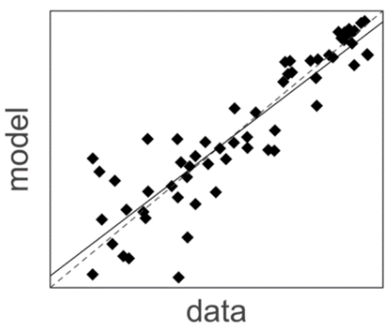
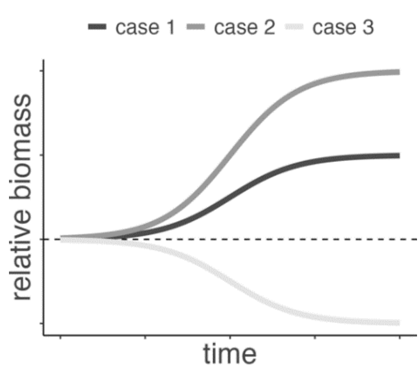
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
The handling of physical components should qualitatively match expectations			
Visual inspection of time-series to evaluate temporal patterns, a combination of error metrics, and/or correlation coefficients (Hipsey et al., 2020).	In both the informal and formal reviews, this aspect was not addressed. However, it should receive attention moving forward, given the interest in exploring the impacts of changing environmental conditions. In Atlantis, temperature is one of the most influential physical components, potentially impacting bioenergetics, movement, recruitment, and assimilation efficiency (Audzijonyte et al., 2017a, 2019). The handling of temperature should therefore be considered in any review. In addition to highly influential physical components, those essential to addressing the research question(s) should also be considered.		Xue et al. (2013).
The handling of water quality and biogeochemistry should qualitatively match expectations			
Visual inspection of time-series to evaluate temporal patterns, a combination of error metrics, and/or correlation coefficients (Hipsey et al., 2020).	In both the informal and formal reviews, this aspect was not addressed. However, it should receive attention moving forward, given the interest in exploring the impacts of changing environmental conditions. Hipsey et al. (2020) summarized pertinent components to consider (e.g., oxygen, nutrients, organic matter, pollutants). Those highly influential in the MEM and/or pertinent to management should be considered. For instance, oxygen is often of focus given it is crucial to nutrient cycling and sediment processes, such as the GOM 'dead zone' (e.g., de Mutsert et al., 2016).		Fennel et al. (2011); Feng et al. (2019).

Table 4 (*continued*)

Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
The handling of group responses to the environment should qualitatively match expectations			
Hipsey et al. (2020) summarized a variety of properties to consider when assessing modeled ecosystem responses to stress, some of which focus on population responses. Assessment is specific to individual cases and the responses being modeled.	The informal review revealed that there is much interest in exploring the impacts of changing environmental conditions on shrimp populations in the GOM. For the formal review, "proof-of-concept" scenarios were presented to showcase features exploring temperature and salinity sensitivity on shrimp movement and temperature sensitivity on shrimp recruitment. While these proof-of-concept results were encouraging, the formal review panel recommended additional development and vetting of the model (at large) before the proof-of-concept scenarios could be formally vetted.		

v) Standards for spatially explicit models

Spatially explicit MEMs are tasked with the additional responsibility of reflecting observed spatio-temporal patterns (Steenbeek et al., 2021). Evaluating these spatial patterns is essential for validating both the MEM itself and its appropriateness for an intended application. Ecosystem processes occur at various spatial scales, while spatially explicit MEMs are constructed for a specific scale, sometimes constrained by the available data.

a. Spatial patterns for biological groups are consistent with data/expectations. Demonstrating that a model properly reflects spatial trends in group biomass is essential to establish reasonable population dynamics. Additional metrics pertinent to modeled groups and/or community structure should be considered if applicable (e.g., predator-prey dynamics, biodiversity).
b. Spatial patterns for modeled fisheries are consistent with data/expectations. Demonstrating that a model properly reflects spatial trends in catch is essential to establish reasonable ecosystem function, and the model's utility as a tool for testing fisheries management scenarios (e.g., marine protected areas).
c. Spatial patterns of environmental components are consistent with data/expectations. Demonstrating that a model properly reflects spatial trends in key environmental components (e.g., temperature, oxygen) is essential to establish a reasonable representation of the environment, and the model's utility as a tool for testing scenarios of changing ocean conditions.

3.3 Account for uncertainty

In conjunction with performance evaluation, uncertainty analysis is a valuable element to include in a review of a MEM (Kaplan and Marshall, 2016). Considering uncertainty can enhance performance evaluation (e.g., comparing a full range of possible outcomes against observations) and demonstrate impacts to model outputs (i.e., prediction capabilities; Steenbeek et al., 2021). Assessing uncertainty can be particularly complicated for MEMs considering the increased parameterization and data demands (Link et al., 2012; Geary et al., 2020; Steenbeek et al., 2021). The GOM Atlantis model, for instance, required about 48 hours of computation time to simulate 50 years of ecosystem dynamics (even more on other machines). Computational limitations pose a major restriction on the assessment of model uncertainty, for which solutions are needed (e.g., a technical remote execution framework that supports the simultaneous execution of multiple simulations; Steenbeek et al., 2021). Below, we outline three methods to consider to account for uncertainty: perturbation analysis on a selection of parameters, model emulation, and bounded scenarios.

- *Perturbation analysis on a selection of parameters*: Model uncertainty is traditionally assessed by analyzing how perturbations to inputs influence model outputs (e.g., sensitivity analysis). A full sensitivity analysis is considered unattainable for models such as Atlantis due to the complexity, run time, and data storage demands (Gaichas et al., 2012). However, it has been shown that sensitivity analysis on a selection of parameters is attainable through perturbation exercises, including model initialization (McGregor et al., 2020), parameters associated with key processes (e.g., Ortega-Cisneros et al., 2017; Sturludottir et al., 2018; Hansen et al., 2019; de Gamiz-Zearra et al., 2024), or parameters often tuned during calibration (e.g., Bracis et al., 2020). Parameters should be selected to focus a model review to areas of uncertainty pertinent to the application.
- *Model emulation*: Model emulation involves developing a computationally efficient surrogate model for a complex model. As long as the emulator outputs are similar to those from the complex model, it provides an efficient means to perform sensitivity analyses for computationally intensive environmental models (Ryan et al., 2018; Aleksankina et al., 2019). Examples include Morzaria-Luna et al. (2018), which demonstrates using a statistical emulator of GOM Atlantis as a means to quantify uncertainty in diet estimates, and Ni et al. (2023), which demonstrates using neural networks to capture key dynamics in an Atlantis model of the Norwegian and Barents Seas. During the formal review, we overviewed the application presented by Morzaria-Luna et al. (2018), and the formal review panel found this methodology capable of providing some estimates of model uncertainty. The growing field of artificial intelligence may provide additional opportunities for model emulation (Heymans et al., 2018; Perryman et al., 2023a).
- *Bounded scenarios*: Despite the merits of the above approaches, significant computational demands still persist. For instance, de Gamiz-Zearra et al. (2024) dedicated an extensive Morris screening approach to explore model sensitivity of the Bay of Biscay Atlantis model, requiring 390 simulations in that dedicated publication. Bounded scenarios offer an approach in lieu of a sensitivity analysis (Kaplan and Marshall, 2016; Audzijonyte et al., 2019; Hansen et al., 2019; Bracis et al., 2020). This entails testing scenarios that represent extreme sets and/or combinations of parameter perturbations that still produce realistic results. This approach limits the number of runs (i.e., computation time and needed storage space) while allowing some estimate of uncertainty. During the formal review, we overviewed bounded scenarios, and the formal review panel found this methodology capable of providing some estimates of model uncertainty. Scenarios can explore either pulse (i.e. temporary) or press (i.e. sustained) parameter perturbations.

Although parameter and initialization uncertainty were of focus during the formal review, uncertainties for MEMs have been generalized into the following additional categories: internal variability, process, structural, and scenario/future (Payne et al., 2016; Geary et al., 2020). Parameter uncertainty is often of focus (Steenbeek et al., 2021) but, depending on the context of the review, other forms of model uncertainty may need to be addressed. For instance, scenario/future uncertainty is likely to dominate under a long-term climate change context (Cheung et al., 2016), and MEMs are therefore being increasingly tested under alternate future scenarios for emissions or ocean physics (Liu et al., 2025; Nilsen et al., 2025). A MEM review could examine scenario/future uncertainty by examining not only the structural pathways simulating future ocean conditions, but also model predictions under future ocean conditions (e.g., scenario analysis; Hodgson et al., 2018). Additionally, structural uncertainty of MEMs is of growing interest, and one novel way of handling this is ensemble modeling. As demonstrated by Spence et al. (2018) in the North Sea, this approach jointly considers structurally dissimilar models, accounting for prior information, fit, and discrepancies between models. A review of multiple MEMs is challenging (SEDAR, 2020), so a MEM review could examine structural uncertainty through general comparisons. For example, during the formal review, we compared shrimp productivities from GOM Atlantis against those from a selection of EwE models (Perryman et al., 2023b).

Table 5: Recommended baseline standards for evaluating the performance of spatially explicit models, including assessment methods, reflections from our experience reviewing a marine ecosystem model (MEM), visual diagnostic example, and literature example.

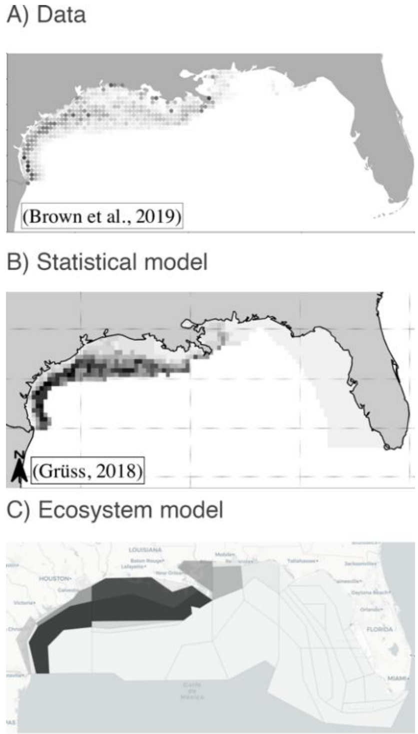
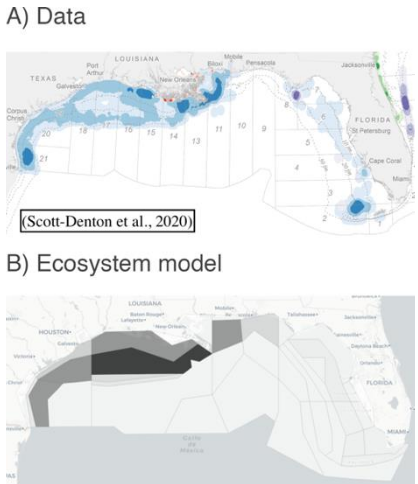
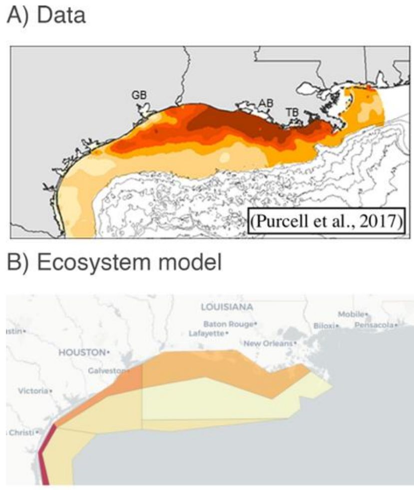
Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Spatial patterns for biological groups are consistent with data/expectations			
Visual inspection of spatial distribution maps, quantitative inspection of a correlation coefficient, and/or quantitative inspection of absolute error. Distribution maps should be evaluated against survey data and/or regional expert opinion, at relevant temporal and spatial ranges.	For many groups in GOM Atlantis, spatial biomass distributions were parameterized based on statistical models (summary provided in Perryman et al., 2023b). The informal review considered Gulf-wide spatial distribution maps for focal species, which resulted in some refinements to the spatial parameterization.		Ortega-Cisneros et al. (2017).
Spatial patterns for modeled fisheries are consistent with data/expectations			
Visual inspection of spatial distribution maps, quantitative inspection of a correlation coefficient, and/or quantitative inspection of absolute error. Distribution maps should be evaluated against survey data and/or regional expert opinion, at relevant temporal and spatial ranges.	The spatial ranges of fleets in GOM Atlantis have been parameterized based on exclusive economic zones (EEZs) and marine protected areas (MPAs) (Ainsworth et al., 2015). The informal review considered Gulf-wide spatial distribution maps of catches from the individual fleets, which resulted in some refinements and updates (Perryman et al., 2023b).		Perryman et al. (2023b).

Table 5 (*continued*)

Assessment methods	Reflections from our experience	Visual diagnostics examples	Examples
Spatial patterns of environmental components are consistent with data/expectations			
Visual inspection of spatial distribution maps, quantitative inspection of a correlation coefficient, and/or quantitative inspection of absolute error. Distribution maps for any environmental components (e.g. bathymetry, temperature, oxygen, species movement responses, etc.) should be evaluated against data and/or regional expert opinion.	In both the informal and formal reviews, this aspect was not addressed. However, it should receive attention moving forward, given the interest in exploring the impacts of changing environmental conditions.		Xue et al. (2013).

4. Digging deeper: supplementary details of the of the Gulf of Mexico Atlantis model review

4.1 Informal review with regional experts

The informal review consisted of regional experts with first-hand knowledge of the GOM, the project focal species (penaeid shrimp), MEMs, and/or EBFM approaches. To focus the GOM Atlantis review, putting the recommendation from Kaplan and Marshall (2016) into practice, the regional experts involved in the informal review helped identify the major interacting species (predators and food sources) for the penaeid shrimp groups. Given that GOM Atlantis is a relatively large model with 91 biological groups and 23 fishing fleets (Ainsworth et al., 2015), identifying these groups set the stage for the discussions during the informal review. The informal nature of the meetings made them highly interactive, fostering open, detailed, and dynamic discussions on the model's diagnostics and realism.

Continuous improvements have been made to GOM Atlantis since its introduction in 2015 (Ainsworth et al., 2015), and the engaging sessions throughout the informal review resulted in additional improvements that progressed the model's realism. A portion of these refinements related to ecological configurations such as group abundances, distributions, and predator-prey linkages, but there was a targeted focus on the configurations of fishing fleets with respect to the spatial dynamics and annual catches. For instance, prior to the informal review, GOM Atlantis handled discarding (catch discarded back to the ecosystem) indirectly for a selection of species. The informal review identified the explicit incorporation of discarding to be a high priority task but acknowledged that it would require extensive effort. As a starting point, the internal review supported the explicit incorporation of dead discarding from fleets pertinent to the focal and key groups: the recreational fisheries (all discarded species), and the commercial shrimp fishery (juvenile red snapper, *Lutjanus campechanus*, and Atlantic croaker, *Micropogonias undulatus*). Dead discards were prioritized due to the influence that the recycled biomass may have on the focal penaeid shrimp groups, however reviewers throughout the entire model review process stressed the importance of explicitly integrating live discards.

Diagnostics developed throughout the informal review were based on the standards for evaluating the performance of end-to-end MEMs recommended by Kaplan and Marshall (2016) (see Section 3.2, all standards under list (i) and standard (ii.a)). Additional diagnostics were developed with regional experts to complement the model improvements accomplished throughout the informal review (see Section 3.2, all standards under list (iii) for both landings and discards, standard (v.a) and standard (v.b). The interactive sessions aided in refining diagnostics (e.g., graphs, charts, tables) that were ultimately used for the performance evaluation in the formal review.

Concurrently with the informal review, we elected to develop an updated technical document of the GOM Atlantis model (Perryman et al., 2023b). This document serves as an overview of the current state of the GOM Atlantis model, including a report of the model improvements and diagnostics resulting from the informal review. Additionally, this document served as an informative guide to the GOM Atlantis model for the formal review.

4.2 Formal review with independent experts

For the formal review, three reviewers were provided by NOAA's CIE program. Following NOAA guidelines (Lynch et al., 2018), the CIE-selected reviewers were not involved in fisheries within the GOM region and had expertise in MEMs, processes affecting marine ecosystems, and/or stock assessment. The formal review panel was supplemented with three regional, external (non-NOAA) experts to provide the CIE reviewers with knowledge on GOM ecology and fisheries. In addition to their regional expertise, the regional external reviewers had expertise in formal model reviews through the Gulf of Mexico Fishery Management Council's Scientific and Statistical Committee (SSC). In the USA, all Fisheries Management Councils use recommendations from their respective SSC to establish fishery management policies or directives. Including reviewers with SSC experience into the formal review was a practical consideration, as uptake of MEM outputs will hinge on managers understanding and then utilizing modeled outcomes.

In conjunction with model performance evaluation, model uncertainty and sensitivity are valuable elements to include in a MEM review (Kaplan and Marshall, 2016; Steenbeek et al., 2021; Kempf et al., 2023). For the formal review, we prepared scenarios to review model sensitivities related to penaeids and their major interacting species, as well as scenarios to review the handling of model uncertainty pertinent to penaeids. Due to limited time and computational resources, we prepared bounded scenarios (see Section 3.3). To review model sensitivities related to penaeids and their major interacting species, we compared the equilibrium responses of shrimp groups in the GOM Atlantis model with those in a selection of GOM-based EwE models and assessed bounded scenarios illustrating the sensitivity of shrimp biomass to perturbations in prey biomass. To review the handling of model uncertainty pertinent to penaeids, we presented an overview of a diet uncertainty analysis by Morzaria-Luna et al. (2018), along with an assessment of bounded scenarios demonstrating the impact of (i) shrimp biomass initialization on focal and key groups and (ii) seagrass biomass on focal and key groups.

Materials were provided to the review panel two weeks prior to the formal review. These included the foundational GOM Atlantis NOAA technical memorandum (Ainsworth et al., 2015), the updated GOM Atlantis NOAA technical memorandum (Perryman et al., 2023b; though unpublished at the time), all published work that used or improved the GOM Atlantis model, an agenda for the formal review, the terms of reference (TOR) for the formal review, a document detailing the roles of the CIE reviewers and rules of the formal review, and an evaluation template for the CIE reviewers.

The meeting structure was based on perspectives from Kaplan and Marshall (2016). The first day consisted of a series of presentations overviewing the motivation and goals of the review, the Atlantis end-to-end modeling framework, the GOM Atlantis model configuration and applications, and updates leading up to the informal review of the GOM Atlantis model. The second day consisted of a series of presentations on the ecology of penaeid shrimp groups, model improvements resulting from the informal review, diagnostics regarding penaeids and their major interacting species groups (see Section 3.1), model sensitivities regarding penaeids and their major interacting species groups, and the handling of model uncertainty pertinent to penaeids. The third day was reserved for final discussions with the formal review panel, review panel deliberation, and report writing. All documentation pertinent to the setup, execution, and results of the GOM Atlantis model review have been made available through a GitHub repository (Perryman, 2024).

5. Conclusions

Formulating fisheries management advice based on the best available scientific information is a commitment being made by numerous countries and intergovernmental organizations, such as the European Commission (European Commission, 2023), ICES (ICES, 2019a), and the Commonwealth Scientific and Industrial Research Organisation (CSIRO; Fulton et al., 2020). This commitment echoes the principles advocated by the Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006 (MSRA) in the United States, highlighting a global consensus on the importance of evidence-based decision-making in sustainable fisheries management. Formal review, as outlined in the US next-generation stock assessment implementation plan by Lynch et al. (2018), is crucial for meeting this mandate. While some aspects of reviewing single-species stock assessment models may be applicable to reviewing MEMs, strategic adaptation is essential based on the MEMs complexity and intended use. Fortunately, in the US, the CIE review system allows for such flexibility. The review process for MEMs presents challenges, but the recommendations provided herein may help guide future reviews. Confronting these challenges is crucial for MEMs to adhere to good modeling practices (Jakeman et al., 2024) and fully realize their potential contributions to EBFM, such as serving as operating models for conducting Management Strategy Evaluation (MSE) (Townsend et al., 2019; Kaplan et al., 2021). Although increasing in frequency (Perryman et al., 2021), MSE guidelines state that operating models for use in MSE should be validated prior to use (ICES, 2019d). In the Southeast US, complex MEMs such as Atlantis are being recognized as promising operating models for conducting MSE because of their ability to capture species interactions and better represent population dynamics (Peterson and Walter, 2023).

Author contributions

H.A.P - Data curation and analysis, visualization, preparation and presentation of review materials, writing - original draft. C.H.A. Project conceptualization, project management, writing - review & editing. M.M. - Project conceptualization, project management, writing - review & editing. I.C.K. - Project conceptualization, preparation of review materials, writing, editing. H.T. - Project guidance, writing – review & editing. S.R.S. - Project guidance, writing – review & editing. M.A.N - Project guidance, writing – review & editing. R.L.S - Preparation of review materials, writing - review & editing. H.C.R - Writing - review & editing.

Funding information

This project was supported by funding from NOAA's Southeast Fisheries Science Center and Fisheries and Office of Science & Technology. Funding for model development was provided by the Florida Restore Act Centers of Excellence Program 8-RCEGR020005-01-02 and RCEGR020428-01-00. I.C.K. acknowledges funding from the NOAA Climate and Fisheries Adaptation Program and the David and Lucile Packard Foundation.

Data availability

All visualizations and data tables were developed using the R statistical software. Scripts and other materials pertinent to this article are freely available on GitHub: https://github.com/hollyannperryman/Gulf-Mexico-Atlantis-Peer-Review.git. The data underlying this article will be shared on a reasonable request to the corresponding author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the regional experts from the National Oceanic and Atmospheric Administration's National Marine Fisheries Service for participating in the informal review. All the participants of the formal review, including the three reviewers provided by the Center of Independent Experts, the three external expert reviewers, the public attendees, and the online participants. Dr. Pierre-Yves Hervann, Dr. Owen Liu, Dr. Jacob Kasper, and Dr. Alberto Rovellini for sharing R-scripts. The scientific results and conclusions, as well as any views and opinions expressed herein, are those of the authors and do not necessarily reflect those of any government agency.

References

Abas, A., Arifin, K., Ali, M. A. M., & Khairil, M. (2023). A systematic literature review on public participation in decision-making for local authority planning: A decade of progress and challenges. Environmental Development, 46, 100853. https://doi.org/10.1016/j.envdev.2023.100853

Ainsworth, C. H., Paris, C. B., Perlin, N., Dornberger, L. N., Patterson, W. F., Chancellor, E., Murawski, S., Hollander, D., Daly, K., Romero, I. C., Coleman, F., & Perryman, H. (2018). Impacts of the Deepwater Horizon oil spill evaluated using an end-to-end ecosystem model. PLOS ONE, 13(1), e0190840. https://doi.org/10.1371/journal.pone.0190840

Ainsworth, C. H., Schirripa, M. J., & Morzaria-Luna, H. N. (2015). An Atlantis Ecosystem Model for the Gulf of Mexico supporting Integrated Ecosystem Assessment (NOAA Technical Memorandum No. NMFS-SEFSC-676; p. 149). http://doi.org/10.7289/V5X63JVH

Ainsworth, C. H., & Walters, C. J. (2015). Ten common mistakes made in Ecopath with Ecosim modelling. Ecological Modelling, 308, 14–17. https://doi.org/10.1016/j.ecolmodel.2015.03.019

Aleksankina, K., Reis, S., Vieno, M., & Heal, M. R. (2019). Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model. Atmospheric Chemistry and Physics, 19(5), 2881–2898. https://doi.org/10.5194/acp-19-2881-2019

Audzijonyte, A., Gorton, R., Kaplan, I., & Fulton, E. A. (2017a). Atlantis User's Guide Part I: General Overview, Physics & Ecology. CSIRO living document.

Audzijonyte, A., Gorton, R., Kaplan, I., & Fulton, E. A. (2017b). Atlantis User's Guide Part II: Socio-Economics. CSIRO living document.

Audzijonyte, A., Pethybridge, H., Porobic, J., Gorton, R., Kaplan, I., & Fulton, E. A. (2019). Atlantis: A spatially explicit end-to-end marine ecosystem model with dynamically integrated physics, ecology and socio-economic modules. Methods in Ecology and Evolution, 10(10), 1814–1819. https://doi.org/10.1111/2041-210X.13272

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., & Andreassian, V. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bracis, C., Lehuta, S., Savina-Rolland, M., Travers-Trolet, M., & Girardin, R. (2020). Improving confidence in complex ecosystem models: The sensitivity analysis of an Atlantis ecosystem model. Ecological Modelling, 431, 109133. https://doi.org/10.1016/j.ecolmodel.2020.109133

Brodziak, J., & Link, J. (2002). Ecosystem-Based Fishery Management: What is it and how can we do it? Bulletin of Marine Science, 70(2), 589–611.

Brown, H., Grüss, A., Hanisko, D., Primrose, J., Rester, J., Rivero, C., Siceloff, L., & Williams, J. (2019). Brown Shrimp In Gulf of Mexico Data Atlas. Stennis Space Center (MS): National Centers for Environmental Information. https://gulfatlas.noaa.gov/

Brown, S. K., Shivlani, M., Die, D., Sampson, D. B., & Ting, T. A. (2006). The Center for Independent Experts: The National External Peer Review Program of NOAA's National Marine Fisheries Service. Fisheries, 31(12), 590–600. https://doi.org/10.1577/1548-8446(2006)31[590:TCFIE]2.0.CO;2

Bryndum-Buchholz, A., Boyce, D., Tittensor, D., Christensen, V., Bianchi, D., & Lotze, H. (2020). Climate-change impacts and fisheries management challenges in the North Atlantic Ocean. Marine Ecology Progress Series, 648, 1–17. https://doi.org/10.3354/meps13438

Cheung, W. W. L., Frölicher, T. L., Asch, R. G., Jones, M. C., Pinsky, M. L., Reygondeau, G., Rodgers, K. B., Rykaczewski, R. R., Sarmiento, J. L., Stock, C., & Watson, J. R. (2016). Building confidence in projections of the responses of living marine resources to climate change. ICES Journal of Marine Science, 73(5), 1283–1296. https://doi.org/10.1093/icesjms/fsv250

Coll, M., & Steenbeek, J. (2017). Standardized ecological indicators to assess aquatic food webs: The ECOIND software plug-in for Ecopath with Ecosim models. Environmental Modelling & Software, 89, 120–130. https://doi.org/10.1016/j.envsoft.2016.12.004

Collie, J. S., Botsford, L. W., Hastings, A., Kaplan, I. C., Largier, J. L., Livingston, P. A., Plagányi, É., Rose, K. A., Wells, B. K., & Werner, F. E. (2016). Ecosystem models for fisheries management: finding the sweet spot. Fish and Fisheries, 17(1), 101–125. https://doi.org/10.1111/faf.12093

Craig, J. K., & Link, J. S. (2023). It is past time to use ecosystem models tactically to support ecosystem-based fisheries management: Case studies using Ecopath with Ecosim in an operational management context. Fish and Fisheries, 24(3), 381–406. https://doi.org/10.1111/faf.12733

de Gamiz-Zearra, A. L., Hansen, C., Corrales, X., & Andonegi, E. (2024). Increasing the reliability of the Bay of Biscay Atlantis model: A sensitivity analysis to parameters perturbations using a Morris screening approach. Ecological Modelling, 488, 110599. https://doi.org/10.1016/j.ecolmodel.2023.110599 de Mutsert, K., Steenbeek, J., Lewis, K., Buszowski, J., Cowan, J. H., & Christensen, V. (2016). Exploring effects of hypoxia on fish and fisheries in the northern Gulf of Mexico using a dynamic spatially explicit ecosystem model. Ecological Modelling, 331, 142–150. https://doi.org/10.1016/j.ecolmodel.2015.10.013

Dornberger, L. N., Montagna, P. A., & Ainsworth, C. H. (2023). Simulating oil-driven abundance changes in benthic marine invertebrates using an ecosystem model. Environmental Pollution, 316, 120450. https://doi.org/10.1016/j.envpol.2022.120450

European Commission. (2023). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions: EU Action Plan: Protecting and restoring marine ecosystems for sustainable and resilient fisheries (No. COM/2023/102 final).

Feng, Y., DiMarco, S. F., Balaguru, K., & Xue, H. (2019). Seasonal and Interannual Variability of Areal Extent of the Gulf of Mexico Hypoxia from a Coupled Physical-Biogeochemical Model: A New Implication for Management Practice. Journal of Geophysical Research: Biogeosciences, 124(7), 1939–1960. https://doi.org/10.1029/2018JG004745

Fennel, K., Hetland, R., Feng, Y., & DiMarco, S. (2011). A coupled physical-biological model of the Northern Gulf of Mexico shelf: model description, validation and analysis of phytoplankton variability. Biogeosciences, 8(7), 1881–1899. https://doi.org/10.5194/bg-8-1881-2011

Fogarty, M. J. (2014). The art of ecosystem-based fishery management. Canadian Journal of Fisheries and Aquatic Sciences, 71(3), 479–490. https://doi.org/10.1139/cjfas-2013-0203

Fulton, E. A., Smith, A. D. M., & Smith, D. C. (2007). Alternative management strategies for southeast Australian Commonwealth Fisheries: stage 2: quantitative management strategy evaluation (p. 380) [Australian Fisheries Management Authority Report]. CSIRO.

Fulton, E. A., van Putten, E. I., Dutra, L. X. C., Melbourne-Thomas, J., Ogier, E., Thomas, L., Murphy, R. P., Butler, I., Ghebrezgabhier, D., Hobday, A. J., & Rayns, N. (2020). Adaptation of fisheries management to climate change Handbook. CSIRO.

Gaichas, S. K., Odell, G., Aydin, K. Y., & Francis, R. C. (2012). Beyond the defaults: functional response parameter space and ecosystem-level fishing thresholds in dynamic food web model simulations. Canadian Journal of Fisheries and Aquatic Sciences, 69(12), 2077–2094. https://doi.org/10.1139/f2012-099

Geary, W. L., Bode, M., Doherty, T. S., Fulton, E. A., Nimmo, D. G., Tulloch, A. I. T., Tulloch, V. J. D., & Ritchie, E. G. (2020). A guide to ecosystem models and their environmental applications. Nature Ecology & Evolution, 4(11), 1459–1471. https://doi.org/10.1038/s41559-020-01298-8

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S. F., Robbins, A. M., … DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. Ecological Modelling, 198(1–2), 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. Ecological Modelling, 221(23), 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Grimm, V., Johnston, A. S. A., Thulke, H.-H., Forbes, V. E., & Thorbek, P. (2020). Three questions to ask before using model outputs for decision support. Nature Communications, 11(1), 4959. https://doi.org/10.1038/s41467-020-17785-2

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., Deangelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M. S., Stillman, R. A., Thiele, J. C., & Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. Journal of Artificial Societies and Social Simulation, 23(2). https://doi.org/10.18564/jasss.4259

Grüss, Arnaud. (2018). Spatial distributions predicted for the paper on monitoring programs of the Gulf of Mexico and the Gulf of Mexico Data Atlas. Distributed by: GRIIDC, Harte Research Institute, Texas A&M University–Corpus Christi. https://doi.org/10.7266/N7GM85QG

Hansen, C., Drinkwater, K. F., Jähkel, A., Fulton, E. A., Gorton, R., & Skern-Mauritzen, M. (2019). Sensitivity of the Norwegian and Barents Sea Atlantis end-to-end ecosystem model to parameter perturbations of key species. PLOS ONE, 14(2), e0210419. https://doi.org/10.1371/journal.pone.0210419

Heymans, J. J., Coll, M., Link, J. S., Mackinson, S., Steenbeek, J., Walters, C., & Christensen, V. (2016). Best practice in Ecopath with Ecosim food-web models for ecosystem-based management. Ecological Modelling, 331, 173–184. https://doi.org/10.1016/j.ecolmodel.2015.12.007

Heymans, S. J. J., Skogen, M., Schrum, C., & Solidoro, C. (2018). Enhancing Europe's capability in end-to-end marine ecosystem modelling for societal benefit. https://doi.org/10.5281/zenodo.3516107

Hipsey, M. R., Gal, G., Arhonditsis, G. B., Carey, C. C., Elliott, J. A., Frassl, M. A., Janse, J. H., De Mora, L., & Robson, B. J. (2020). A system of metrics for the assessment and improvement of aquatic ecosystem models. Environmental Modelling & Software, 128, 104697. https://doi.org/10.1016/j.envsoft.2020.104697

Hodgson, E. E., Kaplan, I. C., Marshall, K. N., Leonard, J., Essington, T. E., Busch, D. S., Fulton, E. A., Harvey, C. J., Hermann, A. J., & McElhany, P. (2018). Consequences of spatially variable ocean acidification in the California Current: Lower pH drives strongest declines in benthic species in southern regions while greatest economic impacts occur in northern regions. Ecological Modelling, 383, 106–117. https://doi.org/10.1016/j.ecolmodel.2018.05.018

Horne, P. J., Kaplan, I. C., Marshall, K. N., Levin, P. S., Harvey, C. J., Hermann, A. J., & Fulton, E. A. (2010). Design and parameterization of a spatially explicit ecosystem model of the central California Current (NOAA Technical Memorandum No. NMFS-NWFSC-104; p. 140). https://repository.library.noaa.gov/view/noaa/3719

Howell, D., Schueller, A. M., Bentley, J. W., Buchheister, A., Chagaris, D., Cieri, M., Drew, K., Lundy, M. G., Pedreschi, D., Reid, D. G., & Townsend, H. (2021). Combining Ecosystem and Single-Species Modeling to Provide Ecosystem-Based Fisheries Management Advice Within Current Management Systems. Frontiers in Marine Science, 7, 607831. https://doi.org/10.3389/fmars.2020.607831

ICES. (2016). Report of the Working Group on Multispecies Assessment Methods (WGSAM) (ICES CM 2015/SSGEPI:20, p. 206). https://doi.org/10.17895/ices.pub.5673

ICES. (2018). Report of the Workshop on Stakeholder Input to, and Parameterization of, Ecosystem and Foodweb Models in the Irish Sea Aimed at a Holistic Approach to the Management of the Main Fish Stocks (WKIrish4) (ICES CM 2017/ACOM:54, p. 35). https://doi.org/10.17895/ices.pub.5427

ICES. (2019a). Advisory Plan (ICES Strategy, p. 18). http://doi.org/10.17895/ices.pub.5468

ICES. (2019b). Report of the Workshop on an Ecosystem-based Approach to Fishery Management for the Irish Sea (WKIrish5) (ICES CM 2018/ACOM:66, p. 55). https://doi.org/10.17895/ices.pub.19291121

ICES. (2019c). Working Group on Multispecies Assessment Methods (WGSAM) (ICES Scientific Reports. 1:91, p. 320). http://doi.org/10.17895/ices.pub.5758

ICES. (2019d). Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2) (ICES Scientific Reports. 1:33, p. 162). http://doi.org/10.17895/ices.pub.5331

ICES. (2023). Working Group on Multispecies Assessment Methods (WGSAM; outputs from 2022 meeting) (ICES Scientific Reports No. 5:12; p. 233). https://doi.org/10.17895/ices.pub.22087292

Kaplan, I. C., Gaichas, S. K., Stawitz, C. C., Lynch, P. D., Marshall, K. N., Deroba, J. J., Masi, M., Brodziak, J. K. T., Aydin, K. Y., Holsman, K., Townsend, H., Tommasi, D., Smith, J. A., Koenigstein, S., Weijerman, M., & Link, J. (2021). Management Strategy Evaluation: Allowing the Light on the Hill to Illuminate More Than One Species. Frontiers in Marine Science, 8, 624355. https://doi.org/10.3389/fmars.2021.624355

Kaplan, I. C., & Marshall, K. N. (2016). A guinea pig's tale: learning to review end-to-end marine ecosystem models for management applications. ICES Journal of Marine Science, 73(7), 1715–1724. https://doi.org/10.1093/icesjms/fsw047

Kempf, A., Spence, M. A., Lehuta, S., Trijoulet, V., Bartolino, V., Villanueva, M. C., & Gaichas, S. K. (2023). Skill assessment of models relevant for the implementation of ecosystem-based fisheries management. Fisheries Research, 268, 106845. https://doi.org/10.1016/j.fishres.2023.106845

Link, J. S. (2010). Adding rigor to ecological network models by evaluating a set of pre-balance diagnostics: A plea for PREBAL. Ecological Modelling, 221(12), 1580–1591. https://doi.org/10.1016/j.ecolmodel.2010.03.012

Link, J. S., Ihde, T. F., Harvey, C. J., Gaichas, S. K., Field, J. C., Brodziak, J. K. T., Townsend, H. M., & Peterman, R. M. (2012). Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. Progress in Oceanography, 102, 102–114. https://doi.org/10.1016/j.pocean.2012.03.008

Link, J. S., Ihde, T. F., Townsend, H. M., Osgood, K. E., Schirripa, M. J., Kobayashi, D. R., Gaichas, S., Field, J. C., Levin, P. S., Aydin, K. Y., & Harvey, C. J. (2010). Report of the 2nd National Ecosystem Modeling Workshop (NEMoW II): bridging the credibility gap dealing with uncertainty in ecosystem models (NOAA Technical Memorandum No. NMFS-F/SPO-102; p. 72). http://spo.nmfs.noaa.gov/tm/

Liu, O. R., Kaplan, I. C., Hernvann, P., Fulton, E. A., Haltuch, M. A., Harvey, C. J., Marshall, K. N., et al. 2025. Climate Change Influences via Species Distribution Shifts and Century-Scale Warming in an End-To-End California Current Ecosystem Model. Global Change Biology, 31: e70021. https://onlinelibrary.wiley.com/doi/10.1111/gcb.70021.

Lynch, P. D., Methot, R. D., & Link, J. S. (2018). Implementing a Next Generation Stock Assessment Enterprise: An Update to the NOAA Fisheries Stock Assessment Improvement Plan (NOAA Technical Memorandum No. NMFS-F/SPO-183; p. 127). http://doi.org/10.7755/TMSPO.183

Marshall, K. N., Koehn, L. E., Levin, P. S., Essington, T. E., & Jensen, O. P. (2019). Inclusion of ecosystem information in US fish stock assessments suggests progress toward ecosystem-based fisheries management. ICES Journal of Marine Science, 76(1), 1–9. https://doi.org/10.1093/icesjms/fsy152

Masi, M. D., Ainsworth, C. H., & Jones, D. L. (2017). Using a Gulf of Mexico Atlantis model to evaluate ecological indicators for sensitivity to fishing mortality and robustness to observation error. Ecological Indicators, 74, 516–525. https://doi.org/10.1016/j.ecolind.2016.11.008

Masi, M. D., Ainsworth, C. H., Kaplan, I. C., & Schirripa, M. J. (2018). Interspecific Interactions May Influence Reef Fish Management Strategies in the Gulf of Mexico. Marine and Coastal Fisheries, 10(1), 24–39. https://doi.org/10.1002/mcf2.10001

McGregor, V. L., Fulton, E. A., & Dunn, M. R. (2020). Addressing initialisation uncertainty for end-to-end ecosystem models: application to the Chatham Rise Atlantis model. PeerJ, 8, e9254. https://doi.org/10.7717/peerj.9254

McGregor, V. L., Horn, P. L., Fulton, E. A., & Dunn, M. R. (2019). From data compilation to model validation: a comprehensive analysis of a full deep-sea ecosystem model of the Chatham Rise. PeerJ, 7, e6517. https://doi.org/10.7717/peerj.6517

Morell, A., Shin, Y.-J., Barrier, N., Travers-Trolet, M., Halouani, G., & Ernande, B. (2023). Bioen-OSMOSE: A bioenergetic marine ecosystem model with physiological response to temperature and oxygen. Progress in Oceanography, 216, 103064. https://doi.org/10.1016/j.pocean.2023.103064

Morzaria-Luna, H., Ainsworth, C., & Scott, R. (2022). Impacts of deep-water spills on mesopelagic communities and implications for the wider pelagic food web. Marine Ecology Progress Series, 681, 37–51. https://doi.org/10.3354/meps13900

Morzaria-Luna, H. N., Ainsworth, C. H., Tarnecki, J. H., & Grüss, A. (2018). Diet composition uncertainty determines impacts on fisheries following an oil spill. Ecosystem Services, 33, 187–198. https://doi.org/10.1016/j.ecoser.2018.05.002

Ni, Y., Sandal, L., Kvamsdal, S., & Hansen, C. (2023). Using feedforward neural networks to represent ecosystem dynamics for bioeconomic analysis. Marine Ecology Progress Series, 716, 1–15. https://doi.org/10.3354/meps14360

Nilsen, I., Hansen, C., and Kaplan, I. C. 2025. A shifting chessboard: Projections of prawn, capelin, mesopelagic fish, zooplankton, and their Nordic and Barents Seas food web under climate change. Progress in Oceanography, 231: 103387. https://doi.org/10.1016/j.pocean.2024.103387.

O'Farrell, H., Grüss, A., Sagarese, S. R., Babcock, E. A., & Rose, K. A. (2017). Ecosystem modeling in the Gulf of Mexico: current status and future needs to address ecosystem-based fisheries management and restoration activities. Reviews in Fish Biology and Fisheries, 27(3), 587–614. https://doi.org/10.1007/s11160-017-9482-1

Olsen, E., Fay, G., Gaichas, S., Gamble, R., Lucey, S., & Link, J. S. (2016). Ecosystem Model Skill Assessment. Yes We Can! PLOS ONE, 11(1), e0146467. https://doi.org/10.1371/journal.pone.0146467

Olsen, E., Kaplan, I. C., Ainsworth, C., Fay, G., Gaichas, S., Gamble, R., Girardin, R., Eide, C. H., Ihde, T. F., Morzaria-Luna, H. N., Johnson, K. F., Savina-Rolland, M., Townsend, H., Weijerman, M., Fulton, E. A., & Link, J. S. (2018). Ocean Futures Under Ocean Acidification, Marine Protection, and Changing Fishing Pressures Explored Using a Worldwide Suite of Ecosystem Models. Frontiers in Marine Science, 5, 64. https://doi.org/10.3389/fmars.2018.00064

Ortega-Cisneros, K., Cochrane, K., & Fulton, E. A. (2017). An Atlantis model of the southern Benguela upwelling system: Validation, sensitivity analysis and insights into ecosystem functioning. Ecological Modelling, 355, 49–63. https://doi.org/10.1016/j.ecolmodel.2017.04.009

Payne, M. R., Barange, M., Cheung, W. W. L., MacKenzie, B. R., Batchelder, H. P., Cormon, X., Eddy, T. D., Fernandes, J. A., Hollowed, A. B., Jones, M. C., Link, J. S., Neubauer, P., Ortiz, I., Queirós, A. M., & Paula, J. R. (2016). Uncertainties in projecting climate-change impacts in marine ecosystems. ICES Journal of Marine Science, 73(5), 1272–1282. https://doi.org/10.1093/icesjms/fsv231

Perryman, H. A. (2024). Gulf-Mexico-Atlantis-Peer-Review. GitHub. https://github.com/hollyannperryman/Gulf-Mexico-Atlantis-Peer-Review.git

Perryman, H. A., Hansen, C., Howell, D., & Olsen, E. (2021). A Review of Applications Evaluating Fisheries Management Scenarios through Marine Ecosystem Models. Reviews in Fisheries Science & Aquaculture, 29(4), 800–835. https://doi.org/10.1080/23308249.2021.1884642

Perryman, H. A., Kaplan, I. C., Blanchard, J. L., Fay, G., Gaichas, S. K., McGregor, V. L., Morzaria-Luna, H. N., Porobic, J., Townsend, H., & Fulton, E. A. (2023a). Atlantis Ecosystem Model Summit 2022: Report from a workshop. Ecological Modelling, 483, 110442. https://doi.org/10.1016/j.ecolmodel.2023.110442

Perryman, H. A., Scott, R. L., Combs-Hintze, B., Repeta, H. C., Vasbinder, K., Masi, M., Kaplan, I., Ainsworth, C. H., Sagarese, S. R., & Nuttall, M. A. (2023b). An Atlantis ecosystem model for the Gulf of Mexico with updates to 2023 (NOAA Technical Memorandum No. NMFS-SEFSC-780; p. 128). doi: 10.25923/y56z-ns92

Perryman, H. A., Tarnecki, J. H., Grüss, A., Babcock, E. A., Sagarese, S. R., Ainsworth, C. H., & Gray DiLeone, A. M. (2020). A revised diet matrix to improve the parameterization of a West Florida Shelf Ecopath model for understanding harmful algal bloom impacts. Ecological Modelling, 416, 108890. https://doi.org/10.1016/j.ecolmodel.2019.108890

Peterson, C. D., & Walter III, J. F. (2023). Southeast Fisheries Science Center Management Strategy Evaluation Strategic Plan (NOAA Technical Memorandum No. NMFS-SEFSC-TM-766; p. 27). https://doi.org/10.25923/khnf-vh41

Pethybridge, H. R., Weijerman, M., Perrymann, H., Audzijonyte, A., Porobic, J., McGregor, V., Girardin, R., Bulman, C., Ortega-Cisneros, K., Sinerchia, M., Hutton, T., Lozano-Montes, H., Mori, M., Novaglio, C., Fay, G., Gorton, R., & Fulton, E. (2019). Calibrating process-based marine ecosystem models: An example case using Atlantis. Ecological Modelling, 412, 108822. https://doi.org/10.1016/j.ecolmodel.2019.108822

Pikitch, E. K., Santora, C., Babcock, E. A., Bakun, A., Bonfil, R., Conover, D. O., Dayton, P., Doukakis, P., Fluharty, D., Heneman, B., Houde, E. D., Link, J., Livingston, P. A., Mangel, M., McAllister, M. K., Pope, J., & Sainsbury, K. J. (2004). Ecosystem-Based Fishery Management. Science, 305(5682), 346–347. https://doi.org/10.1126/science.1098222

Plagányi, É. E. (2007). Models for an ecosystem approach to fisheries (FAO Fisheries Technical Paper No. 447; p. 108). Food and Agriculture Organization of the United Nations. https://www.fao.org/4/a1149e/a1149e.pdf

Planque, B., Aarflot, J. M., Buttay, L., Carroll, J., Fransner, F., Hansen, C., Husson, B., Langangen, Ø., Lindstrøm, U., Pedersen, T., Primicerio, R., Sivel, E., Skogen, M. D., Strombom, E., Stige, L. C., Varpe, Ø., & Yoccoz, N. G. (2022). A standard protocol for describing the evaluation of ecological models. Ecological Modelling, 471, 110059. https://doi.org/10.1016/j.ecolmodel.2022.110059

Porobic Garate, J. (2019). An ecosystem based management framework for the Juan Fernandez Ridge fisheries [Thesis, University Of Tasmania]. https://doi.org/10.25959/100.00031954

Porobic, J. (2022). ReactiveAtlantis: Multiple Reactive HTML tools to help in the calibration of the Atlantis ecosystem model (AEM). R package version 0.0.2.0. https://github.com/Atlantis-Ecosystem-Model/ReactiveAtlantis.git

Purcell, K. M., Craig, J. K., Nance, J. M., Smith, M. D., & Bennear, L. S. (2017). Fleet behavior is responsive to a large-scale environmental disturbance: Hypoxia effects on the spatial dynamics of the northern Gulf of Mexico shrimp fishery. PLOS ONE, 12(8), e0183032. https://doi.org/10.1371/journal.pone.0183032

Reum, J. C. P., Townsend, H., Gaichas, S., Sagarese, S., Kaplan, I. C., & Grüss, A. (2021). It's Not the Destination, It's the Journey: Multispecies Model Ensembles for Ecosystem Approaches to Fisheries Management. Frontiers in Marine Science, 8, 631839. https://doi.org/10.3389/fmars.2021.631839

Rodriguez-Perez, A., Tsikliras, A. C., Gal, G., Steenbeek, J., Falk-Andersson, J., & Heymans, J. J. (2023). Using ecosystem models to inform ecosystem-based fisheries management in Europe: a review of the policy landscape and related stakeholder needs. Frontiers in Marine Science, 10, 1196329. https://doi.org/10.3389/fmars.2023.1196329

Rovellini, A., Punt, A. E., Bryan, M. D., Kaplan, I. C., Dorn, M. W., Aydin, K., Fulton, E. A., Alglave, B., Baker, M. R., Carroll, G., Ferriss, B. E., Haltuch, M. A., Hayes, A. L., Hermann, A. J., Hernvann, P.-Y., Holsman, K. K., Liu, O. R., McHuron, E., Morzaria-Luna, H. N., … Weise, M. T. (2024). Linking climate stressors to ecological processes in ecosystem models, with a case study from the Gulf of Alaska. ICES Journal of Marine Science, fsae002. https://doi.org/10.1093/icesjms/fsae002

Ryan, E., Wild, O., Voulgarakis, A., & Lee, L. (2018). Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. Geoscientific Model Development, 11(8), 3131–3146. https://doi.org/10.5194/gmd-11-3131-2018

Sagarese, S. R., Vaughan, N. R., Walter, J. F., & Karnauskas, M. (2021). Enhancing single-species stock assessments with diverse ecosystem perspectives: a case study for Gulf of Mexico red grouper (Epinephelus morio) and red tides. Canadian Journal of Fisheries and Aquatic Sciences, 78(8), 1168–1180. https://doi.org/10.1139/cjfas-2020-0257

Scott-Denton, E., Cryer, P. F., Duffin, B. V., Duffy, M. R., Gocke, J. P., Harrelson, M. R., Whatley, A. J., & Williams, J. A. (2020). Characterization of the U.S. Gulf of Mexico and South Atlantic Penaeidae and Rock Shrimp (Sicyoniidae) Fisheries through Mandatory Observer Coverage, from 2011 to 2016. Marine Fisheries Review, 82(1–2), 17–47. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00901830&v=2.1&it=r&id=GALE%7CA648333230&sid=googleScholar&linkaccess=abs

SEDAR. (2019). SEDAR 61: Gulf of Mexico Red Grouper (p. 285). https://sedarweb.org/assessments/sedar-61/

SEDAR. (2020). SEDAR 69: Atlantic Menhaden Ecological Reference Points Stock Assessment Report (p. 560). http://sedarweb.org/sedar-69

SEDAR. (2021). SEDAR 72: Gulf of Mexico Gag Grouper (p. 326). https://sedarweb.org/assessments/sedar-72/

Skogen, M., Ji, R., Akimova, A., Daewel, U., Hansen, C., Hjøllo, S., Van Leeuwen, S., Maar, M., Macias, D., Mousing, E., Almroth-Rosell, E., Sailley, S., Spence, M., Troost, T., & Van De Wolfshaar, K. (2021). Disclosing the truth: Are models better than observations? Marine Ecology Progress Series, 680, 7–13. https://doi.org/10.3354/meps13574

Spence, M. A., Blanchard, J. L., Rossberg, A. G., Heath, M. R., Heymans, J. J., Mackinson, S., Serpetti, N., Speirs, D. C., Thorpe, R. B., & Blackwell, P. G. (2018). A general framework for combining ecosystem models. Fish and Fisheries, 19(6), 1031–1042. https://doi.org/10.1111/faf.12310

Steenbeek, J., Buszowski, J., Chagaris, D., Christensen, V., Coll, M., Fulton, E. A., Katsanevakis, S., Lewis, K. A., Mazaris, A. D., Macias, D., De Mutsert, K., Oldford, G., Pennino, M. G., Piroddi, C., Romagnoni, G., Serpetti, N., Shin, Y.-J., Spence, M. A., & Stelzenmüller, V. (2021). Making spatial-temporal marine ecosystem modelling better – A perspective. Environmental Modelling & Software, 145, 105209. https://doi.org/10.1016/j.envsoft.2021.105209

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., Rose, K. A., & Wallhead, P. (2009). Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems, 76(1–2), 4–15. https://doi.org/10.1016/j.jmarsys.2008.03.011

Sturludottir, E., Desjardins, C., Elvarsson, B., Fulton, E. A., Gorton, R., Logemann, K., & Stefansson, G. (2018). End-to-end model of Icelandic waters using the Atlantis framework: Exploring system dynamics and model reliability. Fisheries Research, 207, 9–24. https://doi.org/10.1016/j.fishres.2018.05.026

Tam, J. C., Fay, G., & Link, J. S. (2019). Better Together: The Uses of Ecological and Socio-Economic Indicators With End-to-End Models in Marine Ecosystem Based Management. Frontiers in Marine Science, 6, 560. https://doi.org/10.3389/fmars.2019.00560

Tittensor, D. P., Novaglio, C., Harrison, C. S., Heneghan, R. F., Barrier, N., Bianchi, D., Bopp, L., Bryndum-Buchholz, A., Britten, G. L., Büchner, M., Cheung, W. W. L., Christensen, V., Coll, M., Dunne, J. P., Eddy, T. D., Everett, J. D., Fernandes-Salvador, J. A., Fulton, E. A., Galbraith, E. D., … Blanchard, J. L. (2021). Next-generation ensemble projections reveal higher climate risks for marine ecosystems. Nature Climate Change, 11(11), 973–981. https://doi.org/10.1038/s41558-021-01173-9

Townsend, H., Harvey, C. J., deReynier, Y., Davis, D., Zador, S. G., Gaichas, S., Weijerman, M., Hazen, E. L., & Kaplan, I. C. (2019). Progress on Implementing Ecosystem-Based Fisheries Management in the United States Through the Use of Ecosystem Models and Analysis. Frontiers in Marine Science, 6, 641. https://doi.org/10.3389/fmars.2019.00641

Vilas, D., Buszowski, J., Sagarese, S., Steenbeek, J., Siders, Z., & Chagaris, D. (2023). Evaluating red tide effects on the West Florida Shelf using a spatiotemporal ecosystem modeling framework. Scientific Reports, 13(1), 2541. https://doi.org/10.1038/s41598-023-29327-z

Walters, C. J., Christensen, V., Martell, S. J., & Kitchell, J. F. (2005). Possible ecosystem impacts of applying MSY policies from single-species assessment. ICES Journal of Marine Science, 62(3), 558–568. https://doi.org/10.1016/j.icesjms.2004.12.005

Xue, Z., He, R., Fennel, K., Cai, W.-J., Lohrenz, S., & Hopkinson, C. (2013). Modeling ocean circulation and biogeochemical variability in the Gulf of Mexico. Biogeosciences, 10(11), 7219–7234. https://doi.org/10.5194/bg-10-7219-2013