A skill assessment framework for the Fisheries and Marine Ecosystem Model Intercomparison Project

Nina Rynne[1,2*], Camilla Novaglio[2,3], Julia Blanchard[2,3], Daniele Bianchi[4], Villy Christensen[5,6], Marta Coll[6,7], Jerome Guiet[4], Jeroen Steenbeek[6], Andrea Bryndum-Buchholz[8], Tyler D. Eddy[8], Cheryl Harrison[9], Olivier Maury[10], Kelly Ortega-Cisneros[11], Colleen M. Petrik[12], Derek P. Tittensor[13] & Ryan F. Heneghan[1,14, 15]

[1.] School of Mathematical Sciences, Queensland University of Technology, 4 George St, Brisbane, Queensland, Australia.

[2.] Institute for Marine and Antarctic Studies, University of Tasmania, Castray Esplanade, Hobart, Tasmania, Australia.

[3.] Centre for Marine Socioecology, University of Tasmania, Castray Esplanade, Hobart, Tasmania, Australia.

[4.] Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA.

[5.] Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, Canada.

[6.] Ecopath International Initiative, Barcelona, Spain.

[7.] Institute of Marine Sciences, Barcelona, Spain.

[8.] Centre for Fisheries Ecosystem Research, Fisheries & Marine Institute, Memorial University, St John's, NL, Canada.

[9.] Department of Ocean and Coastal Science and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana, USA.

[10.] IRD, Univ. Montpellier, Ifremer, CNRS, INRAE, MARBEC, Sète France.

[11.] Department of Biological Sciences, University of Cape Town, Cape Town, South Africa.

[12.] Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA.

[13.] Department of Biology, Dalhousie University, Halifax, Canada.

[14.] School of Science, Technology and Engineering, University of the Sunshine Coast, Petrie, Queensland, Australia.

[15.] Australian Rivers Institute, School of Environment and Science, Griffith University, Nathan, QLD, Australia

* Corresponding author: Nina Rynne (nina.rynne@qut.edu.au)

Key Points:

- We developed a standardised skill assessment framework for an ensemble of global marine ecosystem models
- Selected models show agreement with the trajectory of fisheries catch, but exhibit biases compared to observed absolute catch values
- Our framework provides a solid basis to guide global marine ensemble model improvement and increase credibility of ensemble projections

Abstract

Understanding climate change impacts on global marine ecosystems and fisheries requires complex marine ecosystem models, forced by global climate projections, that can robustly detect and project changes. The Fisheries and Marine Ecosystems Model Intercomparison Project (FishMIP) uses an ensemble modelling approach to fill this crucial gap. Yet FishMIP does not have a standardised skill assessment framework to quantify the ability of member models to reproduce past observations and to guide model improvement. In this study, we apply a comprehensive model skill assessment framework to a subset of global FishMIP models that produce historical fisheries catches. We consider a suite of metrics and assess their utility in illustrating the models' ability to reproduce observed fisheries catches. Our findings reveal improvement in model performance at both global and regional (Large Marine Ecosystem) scales from the Coupled Model Intercomparison Project Phase 5 and 6 simulation rounds. Our analysis underscores the importance of employing easily interpretable, relative skill metrics to estimate the capability of models to capture temporal variations, alongside absolute error measures to characterise shifts in the magnitude of these variations between models and across simulation rounds. The skill assessment framework developed and tested here provides a first objective assessment and a baseline of the FishMIP ensemble's skill in reproducing historical catch at the global and regional scale. This assessment can be further improved and systematically applied to test the reliability of FishMIP models across the whole model ensemble from future simulation rounds and include more variables like fish biomass or production.

1 Introduction

Across the world's oceans, marine ecosystems are impacted by humans through fishing, pollution, land use change, and via the accelerating impacts of climate change and ecosystem degradation (Halpern et al., 2008; Hatton et al., 2021). Demand on marine ecosystems for food production is already outpacing human population growth (FAO, 2022), while climate change impacts are expected to perturb marine communities, from individuals to ecosystems (Fulton et al., 2019), driving changes in the availability, resilience, biomass and location of fish stocks (Blanchard et al., 2012; Booth et al., 2017; Cheung et al., 2010; Hollowed et al., 2013; Lotze et al., 2019; Tittensor et al., 2021).

With the growing scope of human impacts on life below water, a range of global Marine Ecosystem Models (MEMs) has been developed by the international research community to help understand and project future change; from simple models based on macroecological scalings to end-to-end models that explicitly represent physical, ecological, and human dynamics; spanning regional systems up to the global ocean. The Fisheries and Marine Ecosystem Model Intercomparison Project (FishMIP; Lotze et al., 2019; Tittensor et al., 2018, 2021; www.fishmip.org) was established in 2013 as part of the broader Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; www.isimip.org) to capitalise on the benefits of bringing such models together into an ensemble. As of today, the FishMIP ensemble comprises no less than nine global and over thirty regional MEMs (Tittensor et al., 2021; Ortega-Cisneros et al., this issue). Individual MEMs are forced by standardised inputs to investigate the influence of environmental conditions and global fishing on ocean biomass and catches while accounting for structural uncertainty across the models (Tittensor et al., 2018, 2021).

Amongst today's most relevant applications of MEMs is quantifying and projecting anthropogenic impacts on marine ecosystems with an overarching goal of informing climate change mitigation and adaptation policy, food security issues, and biodiversity policy (Novaglio et al., 2024). Yet to credibly project anthropogenic impacts on marine ecosystems, the reliability of these MEM projections must be assessed in terms of their skill in reproducing past observations. Skill assessment, broadly, involves comparing model outputs for each member of an ensemble with independent sources of observational data using statistical techniques, and comparing metrics of skill across models (Baumberger et al., 2017; Kubicek et al., 2015; Power, 1993; Stow et al., 2009). Robust skill assessment for ecological models is challenging and still lacks widespread usage. A recent review found that as few as 24% of published ecological modelling studies conducted some form of objective (i.e. metric based) skill assessment (Kubicek et al., 2015), while basic visual comparison is the most commonly used subjective skill assessment method, and is arguably the de facto community standard (Stow et al., 2009). More work is needed to standardise and accelerate the use of model skill assessments to enhance the credibility and reliability of ecosystem model projections for MEMs.

Rigorous model skill assessment needs to address the relevance of models to the scientific or societal question they are addressing (Jakeman et al., 2006; Planque et al., 2022). This includes identifying sets of relevant metrics that help quantify the realism of simulated variables or patterns of importance (Allen & Somerfield, 2009; Bennett et al., 2013; Power, 1993; Stow et al., 2009); and addressing the relevance of the models given technical limitations or the needs of end-users (Hamilton et al., 2019; Kubicek et al., 2015; Steenbeek et al., 2021, this issue).

There are four major challenges that have inhibited the widespread usage of skill assessment for ecological models including MEMs, and especially for cross-model comparison: (1) the absence of a standardised framework, leading to arbitrary and inconsistent choices of important metrics (Geary et al., 2020; Hipsey et al., 2020; Kubicek et al., 2015; Mayer & Butler, 1993; Rykiel, 1996); (2) the need for multiple relevant metrics to assess different aspects of model performance, as relying on a single metric can obscure divergent behaviour or favour models that are highly correlated with a particular set of observations by chance (Bennett et al., 2013; Eyring et al., 2019; Legates & McCabe Jr., 1999; Mayer & Butler, 1993; Power, 1993); (3) credible replication of observations by a model in some regions or for a given time-period does not guarantee performance beyond the calibrated range (Eyring et al., 2019; Hipsey et al., 2020; Hollowed et al., 2013; Refsgaard et al., 2014; Steenbeek et al., 2021; Wagener et al., 2022); (4) the hypothetical nature of future projections makes their comparison to observations unfeasible (Baumberger et al., 2017; Hamilton et al., 2019; Hollowed et al., 2013; Refsgaard et al., 2014).

In the context of fisheries and ecosystem models, these challenges are compounded by limitations in the observational data available to validate MEMs, and the quality of available data. While global and regional catch reconstructions exist (e.g. Pauly & Zeller, 2016; Watson & Tidd, 2018), global observations of fish biomass are lacking. Stock assessments, such as the RAM Legacy Stock Assessment Database (Ricard et al., 2012; www.ramlegacy.org) or recent regional standardised synthesis of biomass observations from trawl surveys (Maureaud et al., 2023) are filling this gap, though they remain limited in spatial coverage, and represent snapshots in time that do not capture variability at seasonal, interannual, or longer timescales.

Although individual skill assessment of FishMIP models has been performed on individual models (Barrier et al., 2023; Blanchard et al., 2012; Carozza et al., 2016, 2017; Cheung et al., 2011; Christensen et al., 2015; Christensen & Walters, 2004; Heneghan et al., 2021; Jennings & Collingridge, 2015; Maury et al., 2007; Maury, 2010; Novaglio et al., 2022; Ortega-Cisneros et al., 2017; Petrik et al., 2019; Sturludottir et al., 2018), these assessments vary from model to model and no standardised skill assessment across the FishMIP model ensemble has taken place yet.

This paper sets the foundations of a skill assessment framework for FishMIP based on the "Concept-State-Process-System" (CSPS) framework created by Hipsey et al. (2020), to build confidence that future predictions are robust and credible. We choose the CSPS framework for its thorough, multi-step process and extensive integration of skill assessment examples in aquatic ecosystem literature.

In choosing and adapting the CSPS framework we attempt to address the first three key challenges (absence of a standardised framework, the need for multiple metrics and model credibility beyond calibrated range) that have hindered the widespread use of model skill assessment in marine ecosystem modelling, and FishMIP in particular. In doing so we: (1) conduct the first standardised model skill assessment of fisheries catch predictions across a subset of FishMIP ensemble members and (2) demonstrate how the CSPS framework can be utilised as the foundation for building context-specific ecosystem skill assessment tools.

We argue that the subsequent uptake of such a framework will improve the credibility and reliability of global MEMs, strengthening their use to inform decision-making. The wider benefits of this case study include illustrating how the framework can be customised for the skill assessment needs of other model intercomparison projects, and therefore further the development of open-access, reliable skill assessment tools for other climate-impact assessment ensembles.

2 Materials and Methods

2.1 Concept-State-Process-System (CSPS) Framework Overview

The CSPS framework categorises measurable elements of ecosystem structure and function across four levels (Hipsey et al., 2020). Level 0 (conceptual assessment) focuses on ensuring that model parameterisations, assumptions and representation of the underlying system are reasonable and derived from a credible scientific basis, and that the level of complexity in the model is appropriate for the questions being asked (Hipsey et al., 2020; Kubicek et al., 2015; Rykiel, 1996). Level 1 (state assessment) is concerned with a model's ability to reflect past observations of measured ecosystem properties. This is generally achieved using metrics that assess goodness-of-fit, which can highlight various mismatches between observations and simulations (Stow et al., 2009). Level 2 (process assessment) and Level 3 (system assessment) explore whether the model is right for the right reasons, i.e., whether the model has captured the important underlying rates of change within the ecosystem, as well as spatial and temporal dynamics that emerge from the model (Hipsey et al., 2020).

Due to the complex task of developing a standardised skill assessment framework for FishMIP models (as detailed in the introduction) and the current lack of a set of quantitative measures to assess FishMIP models' ability to reproduce past trends and patterns, this paper focuses on the implementation of Level 1 model skill assessment. However, we will consider future developments, including the development of methods for Levels 2 and 3 in the context of FishMIP in the Discussion, noting that such methods are still broadly described in the literature and seldom considered when assessing marine ecosystem models.

CSPS addresses three of the four major challenges of MEMs skill assessment: it provides a standardised approach and selection of metrics that can be used across models (addressing the absence of a standardised assessment framework); multiple metrics are used to assess model performance (addressing the need for a suite of metrics to holistically assess model performance); and, finally it provides a framework to assess whether models can replicate ecosystem processes and properties, which is critical for MEMs to provide credible prediction outside their calibrated range (Hipsey et al., 2020; Kubicek et al., 2015; Steenbeek et al., 2021). The identification of emergent processes are context- and model-specific, as the important dynamics to be assessed vary depending on the purpose of the model (Petrik et al., 2022; Novaglio et al., this issue). Identifying relationships between historical simulations and forecasted ocean states through emergent constraints will help address the challenge arising from the hypothetical nature of model projections.

2.2 Metrics for Level 1 FishMIP model assessment

We adapted the CSPS framework to the skill assessment needs of FishMIP in three steps. First, we used the CSPS framework as a benchmark to categorise model skill assessment approaches proposed in other papers (see Table S1). Second, we used best practices and commonly agreed-upon statistical measures to recommend FishMIP-appropriate assessment measures (Table 1). Third, we applied the framework to two structurally contrasting global FishMIP models and assessed their respective ability to reproduce historical fisheries catches at both the global and Large Marine Ecosystem (LME) scale.

We used a synthesis of goodness-of-fit metrics from existing skill assessment literature (Table 1, S2), from which we selected a range of statistical measures for this study (Table 1). Following the advice of Legates & McCabe Jr. (1999) in using both relative and absolute error measures, for our Level 1 assessment we used a visual representation of correlation and bias, and a Taylor diagram plot of standard deviation, Pearson correlation and centred root mean squared error. Alongside this, we calculated 6 independent metrics of model skill assessment (Table 1; Allen & Somerfield, 2009; Bennett et al., 2013; Mayer & Butler, 1993; Stow et al., 2009; Taylor, 2001). These metrics measure: (1) the models' ability to replicate trends over time (Pearson correlation (R)); (2) bias between projections and observations (average error (AE), root mean squared error (RMSE), mean absolute error (MAE)); and (3) a combination of trend and bias (reliability index (RI), and modelling efficiency (MEF)). These skill metrics are detailed in Table 1 and S2. These metrics are calculated for the two FishMIP MEMs, considered here and described in the next section, and are tabulated for comparison.

Table 1. Skill Assessment Metrics. List of skill assessment metrics, their usage, and additional notes. Adapted from Stow et al., (2009). See Table S2 for more information about these metrics.

Name	Type	Ideal Value	Usage	Notes
Correlation (R)	Relative	1	R measures the degree to which simulated and observed catches change together in time. This metric indicates if both variables move in the same direction over time.	R is a relative (or dimensionless) statistic, meaning that correlation is not influenced by the magnitude of the underlying data, therefore values close to 1 can occur even if there is considerable difference in magnitude between the values. Additionally, correlation can be sensitive to outliers if they exist in the data. Relative statistics are comparable across different models or regions.
Average Error (AE)	Absolute	0	AE is the sum of the size of the discrepancies between simulated and observed catch value-pairs. It measures the aggregate bias (or under/overestimation) of simulated catches compared to observations.	A shortcoming of AE is that results close to zero can indicate either a close match or can be a result of positive and negative errors cancelling out. To overcome this, other methods of calculating error can be used instead of, or alongside, AE.
Root Mean Squared Error (RMSE)	Absolute	0	RMSE gives the average distance between predicted and observed catches. It measures the aggregate bias of simulated catches compared to observation Centred RMSE – as reported in a Taylor diagram - is given as a RMSE relative to the standard deviation of observed catches.	RMSE accommodates for the shortcomings of AE as it considers the magnitude, but not the direction, of each discrepancy. As RMSE uses the square of each discrepancy, it is more sensitive to the influence of outliers than either AE or MAE.
Mean Absolute Error (MAE)	Absolute	0	MAE is the sum of the absolute size of the discrepancies between simulated and observed catches. It measures the aggregate bias of the simulated catches compared to observations.	MAE accommodates for the shortcoming of AE, by using the absolute value of the discrepancies. When absolute differences are of a similar magnitude, RMSE and MAE will be approximately equal (Mayer & Butler, 1993)
Reliability Index (RI)	Relative	1	RI is a measure of the average multiplicative factor by which simulated catches differ from observations. Similar to AE, RMSE and MAE, it can be used to measure the bias of the simulations, but as a	RI results are relative making them useful for comparing projections from different models or for different regions

			relative statistic, it can be compared across different models or regions.	
Modelling Efficiency (MEF)	Relative	1	MEF measures the predictive ability of model simulations, relative to the average of the observations in an easily interpretable single statistic.	MEF ∈ (-∞, 1]. A negative result indicates that the observational average is a better predictor than the model projections. Results >0 indicate that the model is a better predictor than the average of the observations.

All analyses were carried out in R-statistics version 4.2.3 (R Core Team, 2023). Statistical metrics were calculated using the R packages *"Metrics"* (Hamner et al., 2018), *"topmodel"* (Buytaert, 2022) and *"qualV"* (Jachner et al., 2007), and the Taylor diagram was plotted using the R package "*openair*" (Carslaw & Ropkins, 2012).

2.3 Marine Ecosystem Models

We obtained model data from two published global MEMs that are members of the FishMIP ensemble and have provided historical simulation outputs of fisheries catches under two FishMIP simulation protocols (using Coupled Model Intercomparison Project (CMIP) Phase 5 and 6 Earth system model forcings, respectively); the BiOeconomic mArine Trophic Size-spectrum model (BOATS; Carozza et al., 2016, 2017) and EcoOcean (Christensen et al., 2015; Coll et al., 2020). These models are a subset of the 9 FishMIP global MEMs (Tittensor et al., 2021), but they are the only two that had historical outputs of fisheries catches across the two simulation rounds at time of publication. Nevertheless, these models capture a significant portion of the spectrum of model complexity across the full FishMIP ensemble, from BOATS which resolves individual organisms by body size alone, to EcoOcean which explicitly incorporates information about thousands of species.

BOATS is a size-structured model that uses broad-scale ecological relationships and individual-level metabolic constraints to calculate the production of fish biomass. It is coupled with an economic module that determines fishing effort and harvest based on the profitability of the exploitation of this biomass given globally homogenous economic boundary conditions (Carozza et al., 2016, 2017). BOATS fish biomass production is driven by water temperature (averaged over the top 75m) and depth-integrated net primary production from Earth System Models (ESMs) and implicitly includes all commercially fished animal biomass from 10 g to 100 kg for three fish groups of increasing asymptotic mass (0.3 kg, 8.5 kg and 100 kg, respectively). By default, in each grid cell of the simulated domain BOATS assumes open-access fishing effort dynamics (Carozza et al., 2016; Tittensor et al., 2018), although it can be forced by observational reconstructions of effort and other social or economic drivers (Scherrer & Galbraith, 2020).

EcoOcean is a combined trophodynamic and species distribution model with a mass-balanced food web model at its core (Christensen et al., 2015; Coll et al., 2020). It uses fishing effort and gear type as forcings, and a gravity model to spatially spread effort across grid cells within LMEs based on expected profitability and fishing costs. Fish prices, used to estimate expected revenue, are model inputs while fishing costs are assumed to be proportional to the grid cell's distance from the nearest coast. Both fish prices and costs are used to calculate fishing effort. The EcoOcean realisation used in this study considers depth-resolved water temperature and depth-integrated small and large phytoplankton biomass as drivers. EcoOcean resolves 51 functional groups, including fish, sharks and rays, invertebrates and mammals, to represent the whole spectrum of marine organisms and integrates explicit information for 3,400 species of marine organisms (Christensen et al., 2015; Coll et al., 2020; Tittensor et al., 2021).

Both MEMs use common simulation protocols, as defined by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; www.isimip.org). We used output from protocols ISIMIP2b and ISIMIP3b (Blanchard et al., 2024; Frieler et al., 2017, 2024; Tittensor et al., 2018, 2021), which used climate forcings from the CMIP5 and CMIP6, respectively.

Total catch output data were provided in a standardised 1° grid cell format monthly. Historical simulations from both MEMs spanned 1971-2005 for ISIMIP2b and 1950-2014 for ISIMIP3b. All outputs from the FishMIP ensemble are available at www.isimip.org/gettingstarted/data-access/. This includes outputs used for the two models presented in this study, except for EcoOcean outputs from ISIMIP3b (available here: 10.5281/zenodo.11081600).

3 Marine Ecosystem Model Forcings and Observational Data

For both ISIMIP2b and 3b, BOATS and EcoOcean were forced with outputs from two Earth System Models (ESMs): Geophysical Fluid Dynamics Laboratories (GFDL) (version ESM2M and ESM4.1 for ISIMIP2b and ISIMIP3b, respectively; Dunne et al., 2012, 2020) and Institut Pierre-Simon Laplace (IPSL) (version CM5A-LR and CM6A-LR for ISIMIP2b and ISIMIP3b, respectively; Boucher et al., 2020; Sepulchre et al., 2020). These ESM simulations forced the FishMIP ensemble models for both ISIMIP simulation rounds (e.g., CMIP5 in Lotze et al., 2019; CMIP5 and CMIP6 in Tittensor et al., 2021).

Two global fishing catch datasets were initially considered to capture the variability and biases from different data reconstruction methodologies. The first, from Watson & Tidd

(2018), covers the historical period 1869-2017 and combines official reconstructed estimates of fisheries catch data, including major discards, from the Food and Agriculture Organisation (FAO) FishStat database along with other publicly available sources (Watson, 2019; Watson & Tidd, 2018; http://dx.doi.org/10.25959/5c522cadbea37). The second, from the Sea Around Us Project (SAUP), covers the historical period 1950-2019, uses FAO-reported landings, Regional Fisheries Management Organisations (RFMOs), expert elicitation and other publicly available sources (Pauly & Zeller, 2016; https://www.seaaroundus.org/data/#/search). However, at the global scale, the difference between these two datasets is small, relative to the divergence between observations and simulations (Figure 1). Therefore, we used only the Watson & Tidd (2018) reconstruction to calculate model performance metrics.

4 Results

4.1 Global scale skill assessment

Globally averaged historical time series of simulated catches were strongly correlated with Watson & Tidd observations for both EcoOcean and BOATS (Figure 1b, 1d). CMIP5-forced correlations were lower for both models compared to CMIP6, (Table 2), especially for BOATS-IPSL (R = 0.47). For CMIP6-forced simulations, correlation coefficients, R, ranged between 0.92 and 0.98, with slightly higher values for BOATS than for EcoOcean (Table 3). Furthermore, bias was substantially lower in CMIP6-forced simulations compared to CMIP5 (Figure 1a vs Figure 1c). Generally, there was greater bias in simulated absolute values of catches compared to observations for BOATS than for EcoOcean across both CMIP5- and 6-forced simulations when using GFDL-forcing (Table 2, Table 3). This result was reversed when using IPSL-forcing, when EcoOcean simulated values showed greater bias. All simulations with EcoOcean generally underestimated catch, while BOATS-GFDL strongly overestimated catch across CMIP5 and 6, whereas the bias of BOATS-IPSL was smaller for both CMIP5 and CMIP6 (Figure 1).

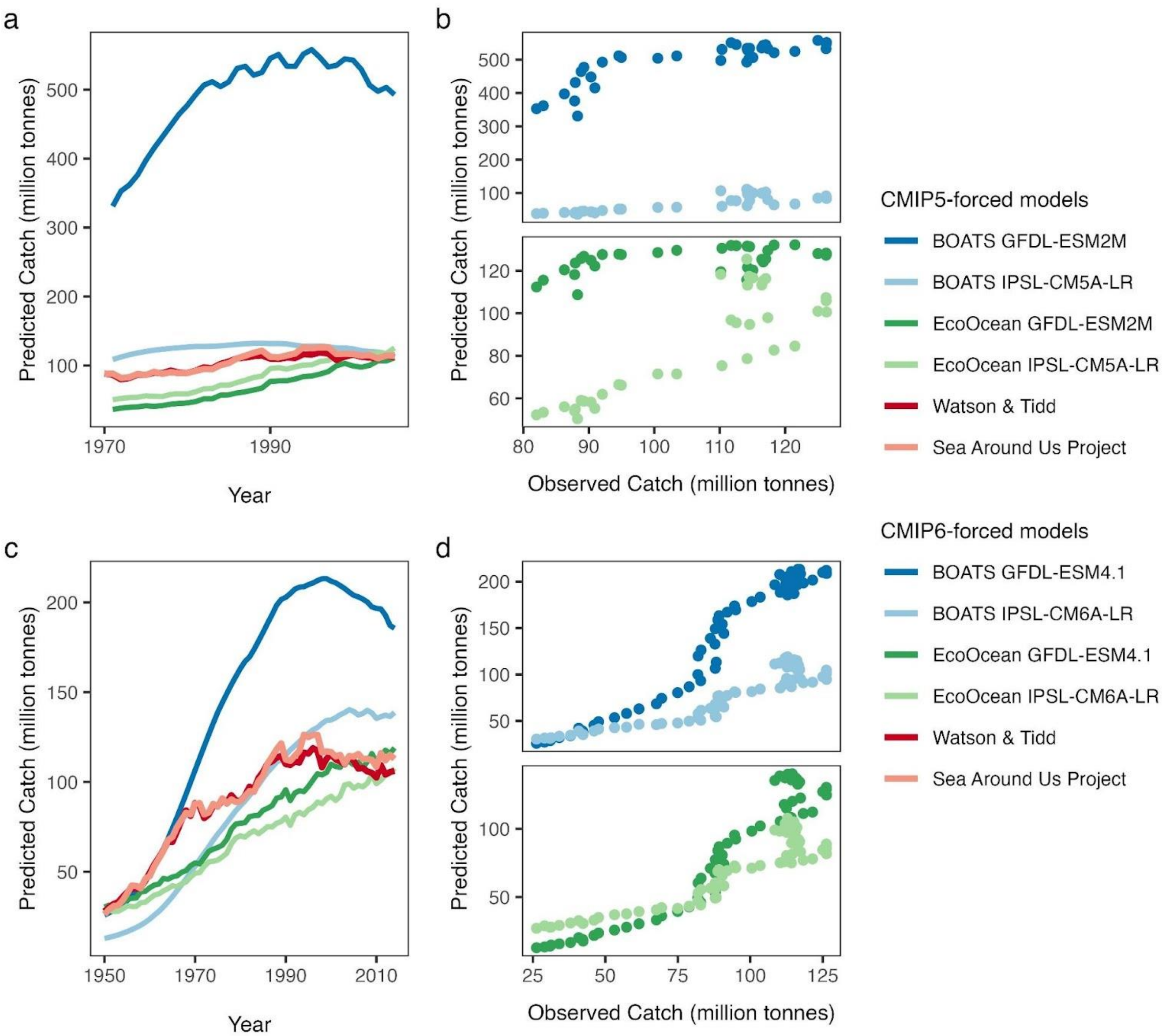

Figure 1. Modelled and observed global fishing catch time series. *Reconstructed observations from Watson & Tidd (2019) and SAUP (2016) and model projected catch for a) CMIP5 from 1971-2005; and c) CMIP6 from 1950-2014. Scatterplot of Watson & Tidd (2019) reconstructed observations vs model predicted catch for b) CMIP5-forced BOATS (top) and EcoOcean (bottom); and d) CMIP6-forced BOATS (top) and EcoOcean (bottom).*

The calculated errors (AE, MAE and RMSE) confirm large discrepancies in the magnitude of simulated and observed catches (Table 2, Table 3). In CMIP6-forced models, the negative result for AE for BOATS-IPSL reflected that simulated catches were lower than observations before ~1980 and higher after ~1990. This led to an AE result closer to zero than other MEM simulations because positive and negative measures cancelled each other out. With the exception of EcoOcean forced by IPSL, all CMIP6-forced models showed improved results for AE, MAE and RMSE compared to CMIP5-forced models (Table 2, 3). RMSE results for all CMIP6-forced models were higher than MAE results, potentially indicating the presence of large outlier values in the simulated catches (Legates & McCabe Jr., 1999).

Table 2. Global forecast skill metrics for fishing catch with CMIP5 forcing. Skill metric performance for six skill metrics using Watson & Tidd (2019) observations: correlation (R), root mean squared error (RMSE; g/m^2), mean absolute error (MAE; g/m^2), average error (AE; g/m^2), reliability index (RI) and modelling efficiency (MEF) for BOATS and EcoOcean models under both ESM forcings. Results in bold are close to ideal results.

	MEM-ESM			
	BOATS		**EcoOcean**	
Skill Metric	**GFDL-ESM2M**	**IPSL-CM5A-LR**	**GFDL-ESM2M**	**IPSL-CM5A-LR**
R	0.84	0.47	0.83	0.86
RMSE	388,936,857	22,825,452	39,586,313	25,900,241
MAE	385,610,855	19,106,219	36,745,703	22,961,393
AE	385,610,855	19,106,219	-36,745,703	-21,598,979
RI	4.64	**1.24**	1.71	1.38
MEF	-755.59	-1.61	-6.84	-2.36

Table 3. Global skill metrics for fishing catch time series with CMIP6 historic forcing. Skill metric performance for six skill metrics using Watson & Tidd (2019) observations: correlation (R), root mean squared error (RMSE; g/m^2), mean absolute error (MAE; g/m^2), average error (AE; g/m^2), reliability index (RI) and modelling efficiency (MEF) for BOATS and EcoOcean models under both ESM forcings. Results in bold are close to ideal results.

	MEM-ESM			
	BOATS		**EcoOcean**	
Skill Metric	**GFDL-ESM4.1**	**IPSL-CM6A-LR**	**GFDL-ESM4.1**	**IPSL-CM6A-LR**
R	**0.98**	**0.95**	**0.92**	**0.92**
RMSE	65,832,789	20,832,218	17,865,414	26,094,844
MAE	54,295,280	17,874,215	14,633,816	23,209,098
AE	53,625,817	-4,835,888	-13,758,893	-23,181,291
RI	**1.57**	**1.51**	**1.26**	**1.41**
MEF	-3.99	**0.5**	**0.633**	**0.216**

RI results showed that all CMIP6-forced models differed from fishing catch by between 1.26-1.57-fold on average compared to observations (Table 3). In contrast, for CMIP5 simulated catches, results differed by up to 4.64-fold (BOATS-IPSL) compared to the observed data (Table 2). This highlighted important improvements in catch estimations between CMIP5- and 6-forced models. In particular, BOATS_GFDL CMIP6-forced runs showed the largest improvement, with an RI of 1.57 for BOATS-GFDL, compared to the CMIP5-forced result of 4.64 (Table 2, 3). Historical catch simulated with CMIP6-IPSL-forced results worsened slightly compared to CMIP5-IPSL-forced runs (Table 3).

For all CMIP5-forced MEM and ESM combinations, modelling efficiency was less than zero, meaning that the average of the observations is more skillful than the models' estimates over the historical period (Table 2). In contrast, modelling efficiency (MEF) was greater than zero for CMIP6-forced BOATS-IPSL, EcoOcean-GFDL and EcoOcean-IPSL (Table 3) indicating that the simulated catches match observed fishing catches more closely than the average of the observations for these simulations. However, modelling efficiency remained negative for CMIP6-forced BOATS-GFDL, albeit greatly improved from CMIP5 (-3.99 versus -755.59) (Table 3).

Taylor diagrams for global catch from BOATS and EcoOcean summarise some of the previous observations. They show an improvement in correlation and bias between CMIP5- and CMIP6-forced simulations for both models however, BOATS-IPSL simulations show an increase in standard deviation from CMIP5- to CMIP6-forced runs (Figure 2).

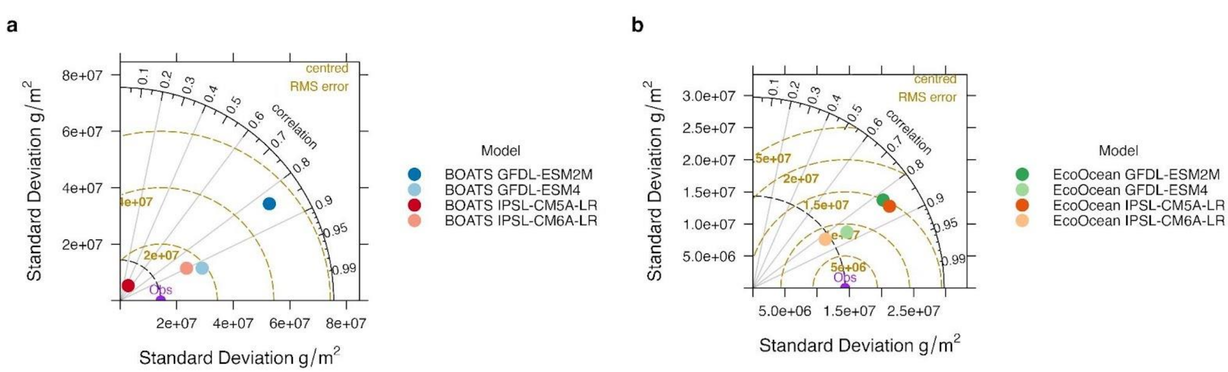

Figure 2. Taylor Diagram for CMIP5 and CMIP6 simulations a) *BOATS model predicted global catch (from 1971-2005) for CMIP6 using ESM IPSL (light red) and GFDL (light blue), and CMIP5 ESM IPSL (dark red) and GFDL (dark blue); and b) EcoOcean model predicted global catch (from 1971-2005) for CMIP6 using ESM IPSL (light orange) and GFDL (light green), and CMIP5 ESM IPSL (dark*

orange) and GFDL (dark green). Plot shows standard deviation, correlation and centred root mean squared error for all 4 MEM-ESM combinations. Watson & Tidd (2019) observed global catch (between 1971-2005) in purple.

4.2 Large Marine Ecosystem scale assessment

Correlations between simulated catches and Watson & Tidd (2019) reconstructed observed fishing catch varied across the LMEs, indicating differing levels of model performance at the regional scale. For all CMIP5-forced models, the median correlation (median across LME-levels) was near zero (Figure 3). Results from CMIP6-forced models showed improvement in the median and interquartile range of correlation results compared to CMIP5-forced models (Figure 3), indicating improved correlation at the LME scale overall.

Geographically, this improvement in correlation from CMIP5 to CMIP6 was evident for both BOATS and EcoOcean, but the degree of improvement (and where this occurs) differs between ESM-forcings (Figure 4). For BOATS, there was an improvement in correlations in CMIP6 compared to CMIP5 across 50 LMEs with both GFDL and IPSL forcing. Similarly, for EcoOcean there was an improvement in correlations across 47 and 46 LMEs for GFDL and IPSL, respectively (Table S3). BOATS showed marked improvements in highly productive LMEs including in the Humboldt Current, Pacific Central-American Coast, Barents Sea, North Brazil Shelf, Patagonian Shelf and Canary Current (Figure 4; Table S3). In contrast, EcoOcean's largest correlation improvements were more randomly dispersed across European, east African and East South American LMEs (Figure 4; Table S3). Negative correlations between observed and modelled catches persist across all simulations in polar regions (Figure 4).

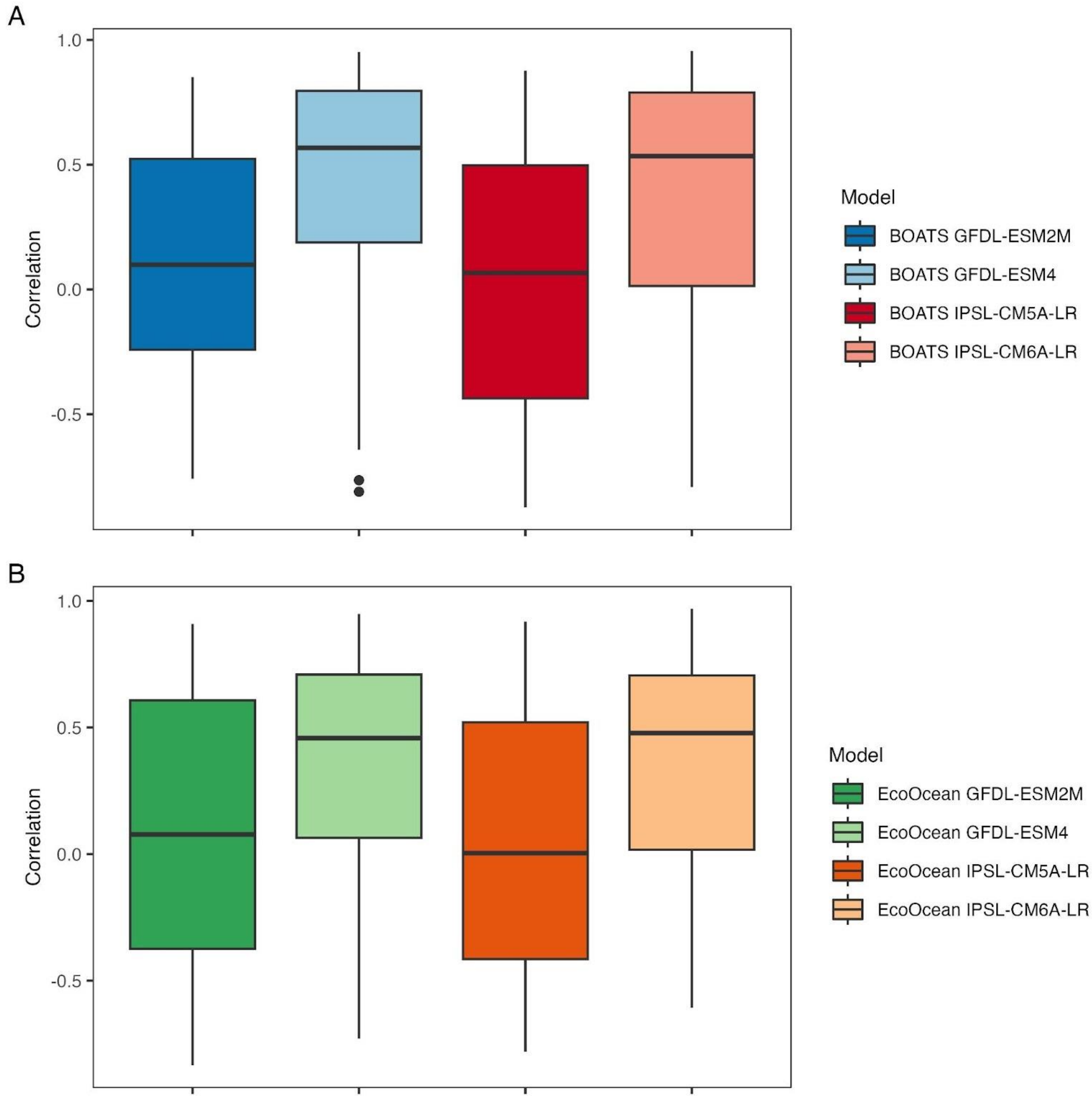

Figure 3. Box plot of Large Marine Ecosystem (LME) level correlation. *a) BOATS model correlation compared to reconstructed observations from Watson & Tidd (2019). CMIP5-forced ESM IPSL (dark red) and GFDL (dark blue), and CMIP6-forced ESM IPSL (light red) and GFDL (light blue); b) EcoOcean model correlation compared to reconstructed observations from Watson & Tidd (2019). CMIP5-forced ESM IPSL (dark orange) and GFDL (dark green), and CMIP6-forced ESM IPSL (light orange) and GFDL (light green).*

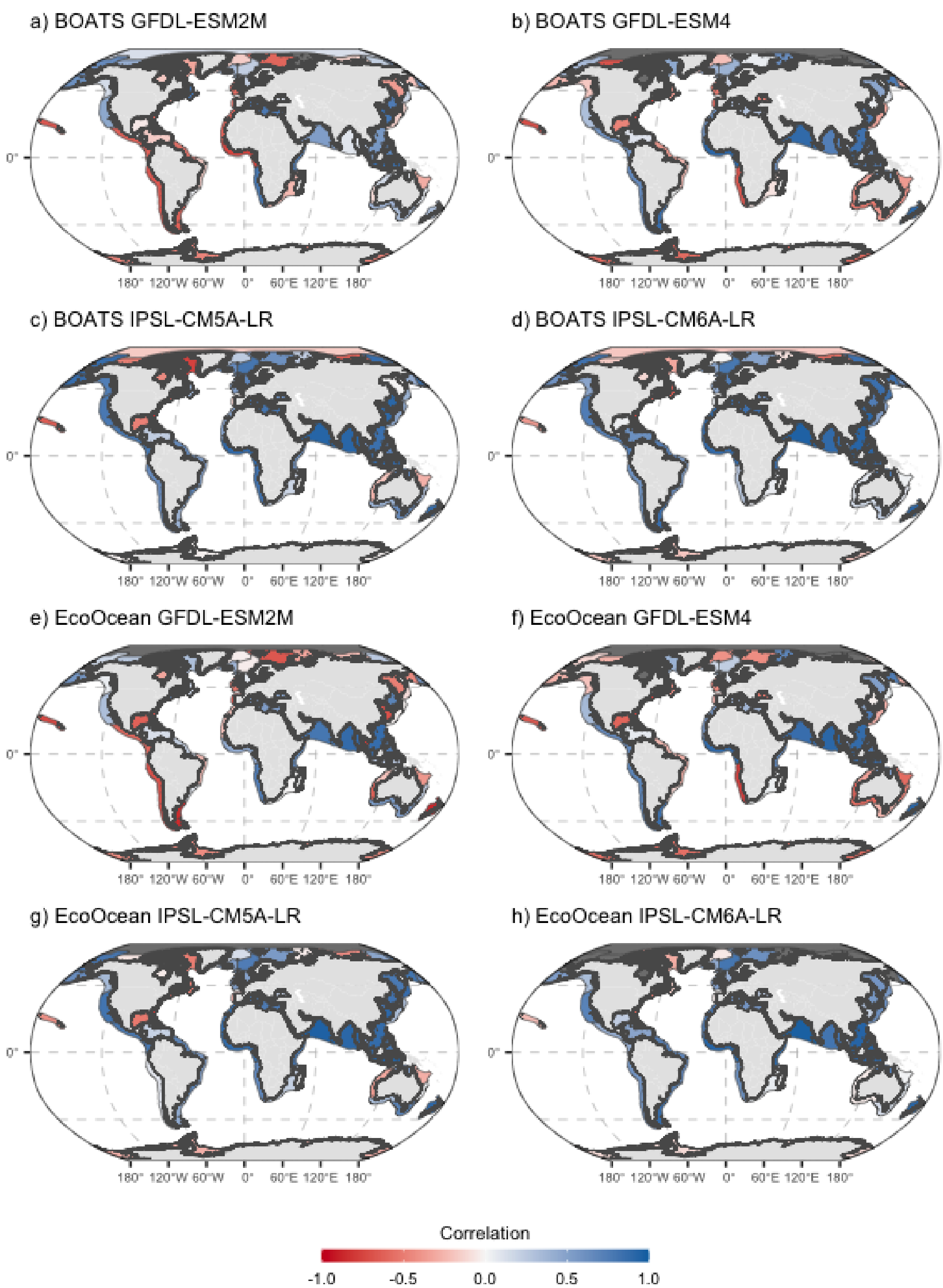

Figure 4. Map of Pearson correlations across the world's LMEs. *BOATS and EcoOcean predictions and reconstructed observations from Watson & Tidd (2019), using CMIP5 from 1971-2005 (a, c, e, g) and CMIP6 from 1950-2014 (b, d, f, h) ESM forcings.*

5 Discussion

Model skill assessment is essential for improving the credibility and reliability of Marine Ecosystem Model (MEM) simulations and for supporting their use as decision-making tools. Until now Fisheries and Marine Ecosystem Model Intercomparison Project (FishMIP) models have generally been assessed in isolation. Here, we adapted and implemented a standardised skill assessment framework to highlight commonalities and discrepancies between modelled and observed historical fish catch and to investigate the usefulness of a range of skill assessment metrics.

Agreement between simulated and observed catches across BOATS and EcoOcean, both in catch time-series variability and absolute catches, is generally higher for CMIP6- than CMIP5-forced models (Figure 1; Table 1, 2). These improvements may be due to changes in the MEMs. For example, EcoOcean has recently undergone substantial restructuring and the upgrades include an expanded food web from 1,400 to over 3,400 explicitly considered individual species, updated functional group representation, and the use of observed historical spatial ranges of species to initialise the model (Coll et al., 2020). This has resulted in an improved understanding and representation of key ecological and fishing dynamics. Between CMIP5 and CMIP6 BOATS' biological formulation and parameters were not changed. However, to improve the match with observed catches, starting effective effort (which then increases through time with improving catchability) was calibrated to align the model's aggregated catch by LMEs with observational reconstructions from Sea Around Us Project (SAUP; Pauly & Zeller, 2016).

While these modifications and – in the case of EcoOcean - a reconsideration of model assumptions, in line with Level 0 (conceptual) assessment, are likely substantial drivers of each model's better performance from CMIP5 to CMIP6, some of the improvement in simulated fishing catches would reflect changes in the two Earth system models (ESMs) that provided input-forcing data to the FishMIP models (Figure S1; Séférian et al., 2020; Tittensor et al., 2021). Across CMIP5 and CMIP6, both ESMs captured observed long-term mean sea-surface temperature (a driver of both models) at the LME scale (Figure S1c, d). Although both models were 0.2-1°C warmer in CMIP6 than CMIP5, they did not show much improvement in resolving net primary production (NPP; a BOATS environmental forcing) at the LME scale (Figure S1a, b). However, across both models, phytoplankton carbon (an EcoOcean forcing) was lower in CMIP6 compared to CMIP5 (Figure S2), which may partly explain why EcoOcean catch bias improved between the two simulations. Ultimately, changes in both MEMs and updated ESM forcings likely contribute to improvements in model skill. Disentangling these drivers would be a fruitful avenue of future research to improve catch and biomass simulations from MEMs.

5.1 Reasons for bias in simulated catches

Bias in fish catch across models are likely driven by a range of factors. For instance, lower trophic level (LTL) biomass and production from ESMs are major drivers of the projected spatial distribution of fish biomass and fisheries catches (Chassot et al., 2010; Heneghan et al., 2021; Kwiatkowski et al., 2020; Laufkötter et al., 2015; Stock et al., 2017; Tagliabue et al., 2021). Thus, discrepancies between observed and modelled LTL variables at the LME scale (Figure S1a, b) will have an impact on MEM fish biomass and therefore catches. In the case of EcoOcean, the one-way forcing of phytoplankton biomass from ESM estimates potentially allows for bias in the estimation of higher trophic level (HTL) biomass, compared to what could be supported by LTL in a two-way coupling. In the case of BOATS, a single energy pathway connects NPP to the accumulation of commercially exploited fish species, while in reality surface pelagic species and bottom living species rely on different food chains, and experience different water temperatures (van Denderen et al., 2018). This may lead to an overestimation of demersal biomass in BOATS, ultimately leading to excess global catches (Guiet et al., 2024). Finally, internal climate variability within the ESMs used here was not calibrated to observed variability. This means that seasonal and annual climate patterns affecting simulated catches from the MEMs will not match observation over these shorter time scales.

The ecological and fishing components of BOATS and EcoOcean, as with all MEMs, are highly simplified representations of the real world that also differ between models. It is important to note that catch reconstructions are also approximations of reality, subject to numerous sources of bias and error. This necessarily results in discrepancies between models, catch reconstructions and actual catches. For example, fishing in BOATS is here determined by globally homogenous and historically constant economic factors, such as constant fish price and fishing costs, and the development of fisheries in each grid cell is driven by the assumption of a historical increase at a constant rate of the technology-driven catchability of fish biomass, which turns initially unprofitable fishing grounds into profitable ones that can be exploited. This assumption is completed with the assumption of open, unregulated fishing access in the world's oceans (Carozza et al., 2016, 2017). These simplifications can capture broad trends in catches at a global level (Galbraith et al., 2017; Guiet et al., 2020), which show a steep increase until the mid-1990s and a later plateau or decline due to overexploitation of fish stocks, and in space a sequential shift of the development of fisheries from cool and productive to warm and unproductive regions (Pauly & Zeller, 2016; Watson & Tidd, 2018). However, they lack important dynamics, which may restrict or change patterns in fishing effort and therefore catches in the real world, particularly at the LME scale considered here.

Finally, other internal issues regarding the way key bio-ecological processes are parameterised can lead to divergence between observed and modelled catch rates, as well as the wide differences between BOATS and EcoOcean historical simulations. To identify specific internal drivers of bias and errors, further experimental attribution simulations would be necessary to separate the impact of individual ecological and fisheries components within each model (Steenbeek et al., this issue). Such experimental studies have already successfully identified key drivers of structural uncertainty in the FishMIP ensemble (Heneghan et al., 2021), and biogeochemical modelling community (Laufkötter et al., 2015).

5.2 Evaluation of metrics for marine model assessment

Our case study highlights how the use of multiple metrics is necessary to obtain a multi-faceted perspective of the credibility and reliability of model projections. While summary statistics correlation (R), reliability index (RI) and modelling efficiency (MEF) provide quick and useful information about model credibility, allow comparison between models or regions, and are generally easy to interpret, some important information about the ecosystem is necessarily lost as these metrics reduce the time-series into a single datum (Bennett et al., 2013; Stow et al., 2009). In addition, these metrics do not assess the ability of the models to capture observed geographical patterns in catches, with this shortcoming highlighting the need for spatial skill assessment tools, such as pattern correlation tools, to be developed and applied in parallel.

Providing too many metrics that measure the same aspect of model skill may instil false confidence in model assessment by unnecessarily replicating the same result (Olsen et al., 2016). Root mean squared error (RMSE), mean absolute error (MAE) and average error (AE) are all slightly different ways of calculating the same measure – the magnitude of the bias between model simulations and observations. Therefore, it is not necessary to use all three bias metrics moving forward (AE, MAE, RMSE). As AE can be affected by positive and negative discrepancies cancelling each other out, and as RMSE is more sensitive to outliers than MAE, we recommend that MAE be the primary metric used to measure bias. On the other hand, evaluating too few metrics, or metrics that only explore one component of model performance can also instil false confidence. For instance, in previous FishMIP syntheses, model outputs were normalised to show only relative changes (Heneghan et al., 2021; Lotze et al., 2019; Tittensor et al., 2021). Normalisation made it possible to generate a more consistent picture of climate change impacts on marine animal biomass, but may have given a false impression of model agreement since it omitted information on discrepancies in absolute biomass across models. Therefore, we argue that it is essential for metrics exploring both absolute and normalised quantities, as used in this case study, to be deployed when assessing MEM performance.

5.3 Future Research

This paper sets out the features of the CSPS framework and conducts a Level 1 skill assessment for two models within the FishMIP ensemble, on simulated catch. We finish by discussing how this process can continue to be improved.

Although 9 global MEMs contribute to the FishMIP ensemble, fisheries catch simulations from only two FishMIP models were available for this assessment. However, we expect that other global and regional models will provide catch outputs as part of the current round of simulations (Blanchard et al., 2024; Frieler et al., 2024) and of ongoing FishMIP efforts to design and implement socioeconomic scenarios that consistently simulate fisheries catch across ecosystem models (Blanchard et al., 2024; Maury et al., this issue).

Individual FishMIP models provide a range of integrated outputs besides total catches analysed here, including catches by functional group (i.e., demersal and pelagic) and by size class (e.g., small, medium and large) (Carozza et al., 2016; Cheung et al., 2011; Coll et al., 2020; Maury, 2010; Maury & Poggiale, 2013; Petrik et al., 2019). The analysis of these outputs can add to the model performance picture and can provide insights into modelled ecosystem structure and function. For example, the collapse of large target species and the increase of smaller species due to predation release (Blanchard et al., 2012; Christensen et al., 2014) can drive fisheries catch, but this process is hidden when considering aggregated biomass and catch outputs. Analysis of these existing outputs is an important next step for FishMIP, and forms part of Level 2 (process) and Level 3 (system) assessment in the CSPS framework (Hipsey et al., 2020). Looking ahead, assessing emergent and system-level relationships between ESM variables and MEM output, or between the internal state variables within the MEMs, also offer considerable potential for enhancing Level 2 and Level 3 assessments of MEM performance (Novaglio et al., this issue). Ultimately, an extensive Level 2 and 3 assessment of the FishMIP ensemble will require models to provide outputs that are not currently part of the CMIP and FishMIP protocols, including primary and secondary production rates, biodiversity turnover, trophic transfer rates or growth rates. Eliciting this information in future simulation protocols is therefore critical, since it will provide scope for in-depth assessment of modelled processes across the FishMIP ensemble.

The current simulation round of FishMIP is focussed on "Detection, Attribution & Evaluation" (ISIMIP3a, www.fishmip.org), and therefore aims to tackle issues such as resolution, coastal processes, and standardisation of fishing inputs across models. To that end, finer scale inputs from ESMs may help the performance of MEMs at the regional scale. There also exists the opportunity to use FishMIP simulations coupled to ESM models forced by reanalysis data (Blanchard et al., 2024) which are constrained by observational products of atmospheric drivers, to calibrate MEMs, or to conduct post-hoc correction of FishMIP outputs (Gómara et al., 2021; Maury et al., this issue). Unlike fully-coupled ESM historical simulations, ocean-only reanalysis-based simulations would have climate oscillations like ENSO cycles occurring at the correct times in history, and thus would hopefully produce more skillful comparisons of time series (e.g. Barrier et al., 2023).

6 Conclusions

Performing model skill assessment on complex end-to-end ecosystem models is an essential, yet challenging task, and there is still considerable progress to be made before model simulations replicate historical observations. MEMs play an important role in developing our understanding of climate change impacts on future fisheries catches and marine ecosystems, and how that might affect global food security (Blanchard et al., 2012, 2017; Booth et al., 2017; Cheung et al., 2010; Cinner et al., 2022; Hollowed et al., 2013). Rigorous ensemble model skill assessment increases confidence in using MEM projections to inform policy, as well as identifying priority areas for future model improvement.

Overall, this case study showed that global fishery catch estimates are well correlated with observed trends over time, but both models show important scale mismatches that require further attention. This exercise provides useful information on the performance of two global models contributing to FishMIP and can be further used to drive model development to improve the reliability of climate impact projections, as well as applied more broadly across the whole suite of FishMIP models to enhance the utility of FishMIP as a whole. We finish with a set of summary recommendations for how FishMIP (and other ensemble model projects) could better integrate model ensemble Level 0-3 skill assessment for future simulation protocols:

1. Level 0: A comprehensive understanding of the underlying assumptions and parameterisations across the model ensemble is essential to understand why MEMs agree or disagree under different conditions. Future protocols targeted at disentangling sources of structural uncertainty across the FishMIP ensemble would concretely improve our understanding of why MEMs behave the way they do. This also includes simulation studies focussed on improving our understanding of the linkages and dependencies between MEMs and the ESMs that force them.

2. Level 1: FishMIP should move beyond exploring only relative change in simulated variables across the model ensemble, to assessing absolute change and variability. This will require using assessment metrics that capture model bias, such as MAE, RI, or MEF.
3. Level 2 and 3: To properly assess the processes and emergent properties of ensemble MEMs, future simulation protocols must require modellers to provide more than just aggregate biomass or catch. At the same time, data products on emergent ecosystem properties such as biomass size-spectra need to be assembled at spatial and temporal resolutions appropriate for comparison with global MEMs.

The CSPS framework provides a solid basis for standardising skill assessment for FishMIP, to which other metrics (e.g., size-based metrics) could be added. The hierarchical structure and focus of each level act as clear guidelines to measure the predictive validity of MEMs. These initial results show that, although we are yet to fully assess the current ensemble of global marine models, we have the tools and knowledge to tackle this task.

Acknowledgments

MC acknowledges support from the Spanish project ProOceans (Plan Estatal de Investigación Científica, Técnica y de Innovación, 2020, PID2020-118097RB-I00) and the the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S) to the Institute of Marine Science (ICM-CSIC). CMP acknowledges support from NOAA grants NA20OAR4310438, NA20OAR4310441, and NA20OAR4310442. KOC acknowledges support from the National Research Foundation of South Africa (grant 136481). OM acknowledges support from the European Union's Horizon 2020 research and innovation program under grant agreement N° 817806.

Open Research

All outputs from the FishMIP ensemble are available at www.isimip.org/gettingstarted/data-access/. This includes outputs used for the two models presented in this study, except for EcoOcean outputs from ISIMIP3b (available here: 10.5281/zenodo.11081600).

The global fishing catch datasets used for observations are available at http://dx.doi.org/10.25959/5c522cadbea37 for Watson & Tidd (2018). The Sea Around Us Project dataset is available here: https://www.seaaroundus.org/data/#/search).

A repository for all R code used to create data visualisations is available on Github here: https://github.com/nina-rynne/FishMIP-modelskill.

References

Allen, J. I., & Somerfield, P. J. (2009). A multivariate approach to model skill assessment. *Journal of Marine Systems*, *76*(1–2), 83–94. https://doi.org/10.1016/J.JMARSYS.2008.05.009

Barrier, N., Lengaigne, M., Rault, J., Person, R., Ethé, C., Aumont, O., & Maury, O. (2023). Mechanisms underlying the epipelagic ecosystem response to ENSO in the equatorial Pacific ocean. *Progress in Oceanography*, *213*, 103002. https://doi.org/10.1016/j.pocean.2023.103002

Baumberger, C., Knutti, R., & Hirsch Hadorn, G. (2017). Building confidence in climate model projections: An analysis of inferences from fit. *Wiley Interdiscip. Rev. Clim. Change*, *8*(3), e454. https://doi.org/10.1002/wcc.454

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

Blanchard, J. L., Jennings, S., Holmes, R., Harle, J., Merino, G., Allen, J. I., Holt, J., Dulvy, N. K., & Barange, M. (2012). Potential consequences of climate change for primary production and fish production in large marine ecosystems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1605), 2979–2989. https://doi.org/10.1098/RSTB.2012.0231

Blanchard, J. L., Novaglio, C., Maury, O., Harrison, C. S., Petrik, C. M., Arcos, L. D. F., Ortega-Cisneros, K., Bryndum-Buchholz, A., Eddy, T., Heneghan, R., Roberts, K. E., Schewe, J., Bianchi, D., Guiet, J., Denderen, D. V., Palacios-Abrantes, J., Liu, X., Stock, C. A. A., Rousseau, Y., … Tittensor, D. (2024). *Detecting, attributing, and projecting global marine ecosystem and fisheries change: FishMIP 2.0*. https://doi.org/10.22541/essoar.170594183.33534487/v1

Blanchard, J. L., Watson, R. A., Fulton, E. A., Cottrell, R. S., Nash, K. L., Bryndum-Buchholz, A., Büchner, M., Carozza, D. A., Cheung, W. W. L., Elliott, J., Davidson, L. N. K., Dulvy, N. K., Dunne, J. P., Eddy, T. D., Galbraith, E., Lotze, H. K., Maury, O., Müller, C., Tittensor, D. P., & Jennings, S. (2017). Linked sustainability challenges and trade-offs among fisheries, aquaculture and agriculture. *Nature Ecology and Evolution*, *1*(9), 1240–1249. https://doi.org/10.1038/s41559-017-0258-8

Booth, D. J., Poloczanska, E., Donelson, J. M., Molinos, J. G., & Burrows, M. (2017). Biodiversity and Climate Change in the Oceans. In *Climate Change Impacts on Fisheries and Aquaculture* (pp. 63–89). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119154051.ch4

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., … Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Buytaert, W. (2022). *topmodel: Implementation of the Hydrological Model TOPMODEL in R* (0.7.4) [Computer software]. https://github.com/ICHydro/topmodel

Carozza, D. A., Bianchi, D., & Galbraith, E. D. (2016). The ecological module of BOATS-1.0: A bioenergetically constrained model of marine upper trophic levels suitable for studies of fisheries and ocean biogeochemistry. *Geoscientific Model Development*, *9*(4), 1545–1565. https://doi.org/10.5194/GMD-9-1545-2016

Carozza, D. A., Bianchi, D., & Galbraith, E. D. (2017). Formulation, General Features and Global Calibration of a Bioenergetically-Constrained Fishery Model. *PLOS ONE*, *12*(1), e0169763. https://doi.org/10.1371/journal.pone.0169763

Carslaw, D., & Ropkins, K. (2012). openair—An R package for air quality data analysis. *ENVIRONMENTAL MODELLING & SOFTWARE*, *27–28*(0), 52–61. [Computer software] https://doi.org/10.1016/j.envsoft.2011.09.008

Chassot, E., Bonhommeau, S., Dulvy, N. K., Mélin, F., Watson, R., Gascuel, D., & Le Pape, O. (2010). Global marine primary production constrains fisheries catches. *Ecology Letters*, *13*(4), 495–505. https://doi.org/10.1111/j.1461-0248.2010.01443.x

Cheung, W. W. L., Dunne, J., Sarmiento, J. L., & Pauly, D. (2011). Integrating ecophysiology and plankton dynamics into projected maximum fisheries catch potential under climate change in the Northeast Atlantic. *ICES Journal of Marine Science*, *68*(6), 1008–1018. https://doi.org/10.1093/icesjms/fsr012

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., Zeller, D., & Pauly, D. (2010). Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Global Change Biology*, *16*(1), 24–35. https://doi.org/10.1111/j.1365-2486.2009.01995.x

Christensen, V., Coll, M., Buszowski, J., Cheung, W. W. L., Frölicher, T., Steenbeek, J., Stock, C. A., Watson, R. A., & Walters, C. J. (2015). The global ocean is an ecosystem: Simulating marine life and fisheries. *Global Ecology and Biogeography*, *24*(5), 507–517. https://doi.org/10.1111/geb.12281

Christensen, V., Coll, M., Piroddi, C., Steenbeek, J., Buszowski, J., & Pauly, D. (2014). A century of fish biomass decline in the ocean. *Marine Ecology Progress Series*, *512*, 155–166. https://doi.org/10.3354/meps10946

Christensen, V., & Walters, C. J. (2004). Ecopath with Ecosim: Methods, capabilities and limitations. *Ecological Modelling*, *172*(2), 109–139. https://doi.org/10.1016/j.ecolmodel.2003.09.003

Cinner, J. E., Caldwell, I. R., Thiault, L., Ben, J., Blanchard, J. L., Coll, M., Diedrich, A., Eddy, T. D., Everett, J. D., Folberth, C., Gascuel, D., Guiet, J., Gurney, G. G., Heneghan, R. F., Jägermeyr, J., Jiddawi, N., Lahari, R., Kuange, J., Liu, W., … Pollnac, R. (2022). Potential impacts of climate change on agriculture and fisheries production in 72 tropical coastal communities. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-30991-4

Coll, M., Steenbeek, J., Pennino, M. G., Buszowski, J., Kaschner, K., Lotze, H. K., Rousseau, Y., Tittensor, D. P., Walters, C., Watson, R. A., & Christensen, V. (2020). Advancing Global Ecological Modeling Capabilities to Simulate Future Trajectories of Change in Marine Ecosystems. *Frontiers in Marine Science*, *7*, 741. https://doi.org/10.3389/FMARS.2020.567877/BIBTEX

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., … Zhao, M. (2020). The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2019MS002015. https://doi.org/10.1029/2019MS002015

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., & Zadeh, N. (2012). GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics. *Journal of Climate*, *25*(19), 6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., … Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change 2019 9:2*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

FAO. (2022). *The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation.* FAO. https://doi.org/10.4060/cc0461en

Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., Geiger, T., Halladay, K., Hurtt, G., Mengel, M., Murakami, D., Ostberg, S., Popp, A., Riva, R., Stevanovic, M., … Yamagata, Y. (2017). Assessing the impacts of 1.5 °C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development*, *10*(12), 4321–4345. https://doi.org/10.5194/gmd-10-4321-2017

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., Otto, C., Reyer, C. P. O., Karger, D. N., Malle, J. T., Treu, S., Menz, C., Blanchard, J. L., Harrison, C. S., Petrik, C. M., Eddy, T. D., Ortega-Cisneros, K., Novaglio, C., Rousseau, Y., … Bechtold, M. (2024). Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model

Intercomparison Project (ISIMIP3a). *Geoscientific Model Development*, *17*(1), 1–51. https://doi.org/10.5194/gmd-17-1-2024

Fulton, E. A., Blanchard, J. L., Melbourne-Thomas, J., Plagányi, É. E., & Tulloch, V. J. D. (2019). Where the Ecological Gaps Remain, a Modelers' Perspective. *Frontiers in Ecology and Evolution*, *7*, 424. https://doi.org/10.3389/FEVO.2019.00424/BIBTEX

Galbraith, E. D., Carozza, D. A., & Bianchi, D. (2017). A coupled human-Earth model perspective on long-term trends in the global marine fishery. *Nature Communications 2017 8:1*, *8*(1), 1–7. https://doi.org/10.1038/ncomms14884

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 2020 4:11*, *4*(11), 1459–1471. https://doi.org/10.1038/s41559-020-01298-8

Gómara, I., Rodríguez-Fonseca, B., Mohino, E., Losada, T., Polo, I., & Coll, M. (2021). Skillful prediction of tropical Pacific fisheries provided by Atlantic Niños. *Environmental Research Letters*, *16*(5), 054066. https://doi.org/10.1088/1748-9326/abfa4d

Guiet, J., Bianchi, D., Scherrer, K., Heneghan, R., & Galbraith, E. D. (2024). *Small fish biomass limits the catch potential in the High Seas*. https://doi.org/10.22541/au.170967563.32290483/v1

Guiet, J., Galbraith, E. D., Bianchi, D., & Cheung, W. W. L. (2020). Bioenergetic influence on the historical development and decline of industrial fisheries. *ICES Journal of Marine Science*, *77*(5), 1854–1863. https://doi.org/10.1093/icesjms/fsaa044

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., Casey, K. S., Ebert, C., Fox, H. E., Fujita, R., Heinemann, D., Lenihan, H. S., Madin, E. M.

P., Perry, M. T., Selig, E. R., Spalding, M., Steneck, R., & Watson, R. (2008). A Global Map of Human Impact on Marine Ecosystems. *Science*, *319*(5865), 948–952. https://doi.org/10.1126/science.1149345

Hamilton, S. H., Fu, B., Guillaume, J. H. A., Badham, J., Elsawah, S., Gober, P., Hunt, R. J., Iwanaga, T., Jakeman, A. J., Ames, D. P., Curtis, A., Hill, M. C., Pierce, S. A., & Zare, F. (2019). A framework for characterising and evaluating the effectiveness of environmental modelling. *Environmental Modelling & Software*, *118*, 83–98. https://doi.org/10.1016/J.ENVSOFT.2019.04.008

Hamner, B., Frasco, M., & LeDell, E. (2018). *Metrics: Evaluation Metrics for Machine Learning* (0.1.4) [Computer software]. https://CRAN.R-project.org/package=Metrics

Hatton, I. A., Heneghan, R. F., Bar-On, Y. M., & Galbraith, E. D. (2021). The global ocean size spectrum from bacteria to whales. *Science Advances*, *7*(46), 3732. https://doi.org/10.1126/SCIADV.ABH3732/SUPPL_FILE/SCIADV.ABH3732_DATA_S1.ZIP

Heneghan, R. F., Galbraith, E., Blanchard, J. L., Harrison, C., Barrier, N., Bulman, C., Cheung, W., Coll, M., Eddy, T. D., Erauskin-Extramiana, M., Everett, J. D., Fernandes-Salvador, J. A., Gascuel, D., Guiet, J., Maury, O., Palacios-Abrantes, J., Petrik, C. M., du Pontavice, H., Richardson, A. J., … Tittensor, D. P. (2021). Disentangling diverse responses to climate change among global marine ecosystem models. *Progress in Oceanography*, *198*, 102659. https://doi.org/10.1016/J.POCEAN.2021.102659

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

Hollowed, A. B., Barange, M., Beamish, R. J., Brander, K., Cochrane, K., Drinkwater, K., Foreman, M. G. G., Hare, J. A., Holt, J., Ito, S. I., Kim, S., King, J. R., Loeng, H., Mackenzie, B. R., Mueter, F. J., Okey, T. A., Peck, M. A., Radchenko, V. I., Rice, J. C., … Yamanaka, Y. (2013). Projected impacts of climate change on marine fish and fisheries. *ICES Journal of Marine Science*, *70*(5), 1023–1037. https://doi.org/10.1093/ICESJMS/FST081

Jachner, S., Boogaart, G., & Petzoldt, T. (2007). Statistical Methods for the Qualitative Assessment of Dynamic Models with Time Delay (R Package qualV). *Journal of Statistical Software*, *22*(8), 1–30.[Computer software] https://doi.org/10.18637/jss.v022.i08

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614. https://doi.org/10.1016/J.ENVSOFT.2006.01.004

Jennings, S., & Collingridge, K. (2015). Predicting Consumer Biomass, Size-Structure, Production, Catch Potential, Responses to Fishing and Associated Uncertainties in the World's Marine Ecosystems. *PLOS ONE*, *10*(7), e0133794. https://doi.org/10.1371/journal.pone.0133794

Kubicek, A., Jopp, F., Breckling, B., Lange, C., & Reuter, H. (2015). Context-oriented model validation of individual-based models in ecology: A hierarchically structured approach to validate qualitative, compositional and quantitative characteristics.

Ecological Complexity, *22*, 178–191. https://doi.org/10.1016/J.ECOCOM.2015.03.005

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., … Ziehn, T. (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, *17*(13), 3439–3470. https://doi.org/10.5194/bg-17-3439-2020

Laufkötter, C., Vogt, M., Gruber, N., Aita-Noguchi, M., Aumont, O., Bopp, L., Buitenhuis, E., Doney, S. C., Dunne, J., Hashioka, T., Hauck, J., Hirata, T., John, J., Le Quéré, C., Lima, I. D., Nakano, H., Seferian, R., Totterdell, I., Vichi, M., & Völker, C. (2015). Drivers and uncertainties of future global marine primary production in marine ecosystem models. *Biogeosciences*, *12*(23), 6955–6984. https://doi.org/10.5194/bg-12-6955-2015

Legates, D. R., & McCabe Jr., G. J. (1999). Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Lotze, H. K., Tittensor, D. P., Bryndum-Buchholz, A., Eddy, T. D., Cheung, W. W. L., Galbraith, E. D., Barange, M., Barrier, N., Bianchi, D., Blanchard, J. L., Bopp, L., Büchner, M., Bulman, C. M., Carozza, D. A., Christensen, V., Coll, M., Dunne, J. P., Fulton, E. A., Jennings, S., … Worm, B. (2019). Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. *Proceedings of the*

National Academy of Sciences of the United States of America, *116*(26), 12907–12912. https://doi.org/10.1073/PNAS.1900194116/SUPPL_FILE/PNAS.1900194116.SAPP.PDF

Maureaud, A., Kitchel, Z., Fredston, A., Guralnick, R., Abrantes, J. P., Palomares, D., Pinsky, M., Shackell, N., Thorson, J., Merigot, B., & Consortium, F. (2023). *FISHGLOB: A collaborative infrastructure for marine science and management*. OSF. https://doi.org/10.31219/osf.io/mh46b

Maury, O. (2010). An overview of APECOSM, a spatialized mass balanced "Apex Predators ECOSystem Model" to study physiologically structured tuna population dynamics in their ecosystem. *Progress in Oceanography*, *84*(1–2), 113–117. https://doi.org/10.1016/J.POCEAN.2009.09.013

Maury, O., & Poggiale, J.-C. (2013). From individuals to populations to communities: A dynamic energy budget model of marine ecosystem size-spectrum including life history diversity. *Journal of Theoretical Biology*, *324*, 52–71. https://doi.org/10.1016/j.jtbi.2013.01.018

Maury, O., Shin, Y.-J., Faugeras, B., Ben Ari, T., & Marsac, F. (2007). Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 2: Simulations. *Progress in Oceanography*, *74*(4), 500–514. https://doi.org/10.1016/j.pocean.2007.05.001

Maury et al. (this issue). The Ocean System Pathways: A new scenario framework to investigate the future of ocean ecosystems and fisheries. In Submission.

Mayer, D. G., & Butler, D. G. (1993). Statistical validation. *Ecological Modelling*, *68*(1–2), 21–32. https://doi.org/10.1016/0304-3800(93)90105-2

Novaglio, C., Blanchard, J. L., Plank, M. J., van Putten, E. I., Audzijonyte, A., Porobic, J., & Fulton, E. A. (2022). Exploring trade-offs in mixed fisheries by integrating fleet dynamics into multispecies size-spectrum models. *Journal of Applied Ecology*, *59*(3), 715–728. https://doi.org/10.1111/1365-2664.14086

Novaglio, C., Bryndum-Buchholz, A., Tittensor, D., Eddy, T. D., Lotze, H. K., Harrison, C. S., Heneghan, R., Maury, O., Ortega-Cisneros, K., Petrik, C. M., Roberts, K. E., & Blanchard, J. L. (2024). *The Past and Future of the Fisheries and Marine Ecosystem Model Intercomparison Project*. ESS Open Archive. https://doi.org/10.22541/essoar.170542252.20348236/v1

Novaglio et al. (this issue). Using ecological theory to evaluate and reduce uncertainties of a marine ecosystem model ensemble. In submission.

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

Ortega-Cisneros, K., Cochrane, K., & Fulton, E. (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

Ortega-Cisneros K., Fierros Arcos D., Novaglio C., Eddy T., Coll M., Fulton E., Oliveros-Ramos R. *et al.* Submitted. An integrated workflow for the detection and attribution of marine ecosystem change, from global to regional scales. *Earth's Future*. This Issue MS#2024EF004826

Pauly, D., & Zeller, D. (2016). Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature Communications*, *7*. https://doi.org/10.1038/NCOMMS10244

Petrik, C. M., Luo, J. Y., Heneghan, R. F., Everett, J. D., Harrison, C. S., & Richardson, A. J. (2022). Assessment and Constraint of Mesozooplankton in CMIP6 Earth System Models. *Global Biogeochemical Cycles*, *36*(11), e2022GB007367. https://doi.org/10.1029/2022GB007367

Petrik, C. M., Stock, C. A., Andersen, K. H., van Denderen, P. D., & Watson, J. R. (2019). Bottom-up drivers of global patterns of demersal, forage, and pelagic fishes. *Progress in Oceanography*, *176*, 102124. https://doi.org/10.1016/j.pocean.2019.102124

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

Power, M. (1993). The predictive validation of ecological and environmental models. *Ecological Modelling*, *68*(1–2), 33–50. https://doi.org/10.1016/0304-3800(93)90106-3

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P., & Christensen, J. H. (2014). A framework for testing the ability of models to project climate change and its impacts. *Climatic Change*, *122*(1–2), 271–282. https://doi.org/10.1007/S10584-013-0990-2/FIGURES/3

Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2012). Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries*, *13*(4), 380–398. https://doi.org/10.1111/j.1467-2979.2011.00435.x

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, *90*(3), 229–244. https://doi.org/10.1016/0304-3800(95)00152-2

Scherrer, K., & Galbraith, E. (2020). Regulation strength and technology creep play key roles in global long-term projections of wild capture fisheries. *ICES Journal of Marine Science*, *77*(7–8), 2518–2528. https://doi.org/10.1093/icesjms/fsaa109

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., John, J. G., Li, H., Long, M. C., Luo, J. Y., Nakano, H., Romanou, A., Schwinger, J., … Yamamoto, A. (2020). Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports*, *6*(3), 95–119. https://doi.org/10.1007/s40641-020-00160-0

Sepulchre, P., Caubel, A., Ladant, J.-B., Bopp, L., Boucher, O., Braconnot, P., Brockmann, P., Cozic, A., Donnadieu, Y., Dufresne, J.-L., Estella-Perez, V., Ethé, C., Fluteau, F., Foujols, M.-A., Gastineau, G., Ghattas, J., Hauglustaine, D., Hourdin, F., Kageyama, M., … Tardif, D. (2020). IPSL-CM5A2 – an Earth system model designed for multi-millennial climate simulations. *Geoscientific Model Development*, *13*(7), 3011–3053. https://doi.org/10.5194/gmd-13-3011-2020

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

Steenbeek et al. (this issue). Making ecosystem modelling operational - a novel distributed execution framework to systematically explore ecological responses to divergent climate trajectories. Paper #2023EF004295.

Stock, C. A., John, J. G., Rykaczewski, R. R., Asch, R. G., Cheung, W. W. L., Dunne, J. P., Friedland, K. D., Lam, V. W. Y., Sarmiento, J. L., & Watson, R. A. (2017). Reconciling fisheries catch and ocean productivity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), E1441–E1449. https://doi.org/10.1073/PNAS.1610238114

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

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., & Vialard, J. (2021). Persistent Uncertainties in Ocean Net Primary Production Climate Change Projections at Regional Scales Raise Challenges for Assessing Impacts on Ecosystem Services. *Frontiers in Climate*, *3*. https://www.frontiersin.org/articles/10.3389/fclim.2021.738224

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Tittensor, D. P., Eddy, T. D., Lotze, H. K., Galbraith, E. D., Cheung, W., Barange, M., Blanchard, J. L., Bopp, L., Bryndum-Buchholz, A., Büchner, M., Bulman, C., Carozza, D. A., Christensen, V., Coll, M., Dunne, J. P., Fernandes, J. A., Fulton, E. A., Hobday, A. J., Huber, V., … Walker, N. D. (2018). A protocol for the intercomparison of marine fishery and ecosystem models: Fish-MIP v1.0. *Geoscientific Model Development*, *11*(4), 1421–1442. https://doi.org/10.5194/GMD-11-1421-2018

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 2021 11:11*, *11*(11), 973–981. https://doi.org/10.1038/s41558-021-01173-9

van Denderen, P. D., Lindegren, M., MacKenzie, B. R., Watson, R. A., & Andersen, K. H. (2018). Global patterns in marine predatory fish. *Nature Ecology & Evolution*, *2*(1), Article 1. https://doi.org/10.1038/s41559-017-0388-z

Wagener, T., Reinecke, R., & Pianosi, F. (2022). On the evaluation of climate change impact models. *WIREs Climate Change*, *13*(3), e772. https://doi.org/10.1002/wcc.772

Watson, R. (2019). *Global Fisheries Landings V4.0* (4.0) [dataset]. Institute for Marine and Antarctic Studies (IMAS), University of Tasmania (UTAS). https://doi.org/10.25959/5C522CADBEA37

Watson, R., & Tidd, A. (2018). Mapping nearly a century and a half of global marine fishing: 1869–2015. *Marine Policy*, *93*, 171–177. https://doi.org/10.1016/j.marpol.2018.04.023