Current Research in Environmental Sustainability 9 (2025) 100273

Contents lists available at ScienceDirect

Current Research in Environmental Sustainability

journal homepage: www.sciencedirect.com/journal/current-research-in-environmental-sustainability

A multi-dimensional approach to improve validation practices for qualitative models of marine social-ecological systems

Bruno Oliveira [a,*], Nuno V. Álvaro [b], Furqan Asif [c], Andrea Z. Botelho [d,e], João Canning-Clode [f,g], Daniela Casimiro [d,e], Candelaria Cecilia-Ruano [h], Catherine Chambers [i], Ana C. Costa [d,e], Ana Dinis [f,g], Jesús P. García [h], Ricardo Haroun [h], Unn Laksá [j], Gustavo M. Martins [d,e], Alexander H. McGrath [k], Caterina Mintrone [k,l], Mirjam Carlsdóttir Olsen [j], Manuela I. Parente [d,e], Paola Parretti [d,e], Sarai Pouso [a], Chiara Ravaglioli [k,l], Maria A.M. Ventura [d,e], Talea Weissang [m], Sandra Rybicki [n], Pamela J. Woods [n], Anna H. Olafsdottir [n], Angel Borja [a]

[a] *AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), 20010 Pasaia, Spain*
[b] *Universidade dos Açores, Instituto de Investigação e Tecnologias Agrárias e do Ambiente Campus de Angra do Heroísmo Rua Capitão João d'Ávlia, Pico da Urze PT, 9700-042 Angra do Heroísmo, Portugal*
[c] *Centre for Blue Governance, Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark*
[d] *Faculty of Sciences and Technology, University of the Azores, 9500-321 Ponta Delgada, Portugal*
[e] *CIBIO, Research Centre in Biodiversity and Genetic Resources, InBio Associate Laboratory, BIOPOLIS Program in Genomics, Biodiversity and Land Planning; UNESCO Chair – Land Within Sea: Biodiversity & Sustainability in Atlantic Islands, University of the Azores, 9500-321 Ponta Delgada, Portugal*
[f] *MARE- Marine and Environmental Sciences Centre /ARNET - Aquatic Research Network, Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI), Funchal, Portugal*
[g] *Faculty of Life Sciences, Universidade da Madeira, Funchal, Portugal.*
[h] *ECOAQUA University Institute, University of Las Palmas de Gran Canaria (ULPGC), Marine Science and Technology Park, Ctra. de Taliarte s/n, 35200, Telde, Las Palmas, Spain*
[i] *Stefansson Arctic Institute, Borgir við Norðursloð, 400 Akureyri, Iceland*
[j] *Sjókovin - Blue Resource, Bryggjan 6, Leirvík, Faroe Islands*
[k] *Department of Biology, University of Pisa, Via Derna 1, Pisa, Italy*
[l] *CoNISMa, Consorzio Nazionale Interuniversitario per le Scienze del Mare, Piazzale Flaminio, 9, 00196 Rome, Italy*
[m] *WWF, Svanevej 12, 2400 Copenhagen, NW, Denmark*
[n] *Marine and Freshwater Research Institute, Fornubúðum 5, 220 Hafnarfjörður, Iceland.*

ARTICLE INFO

Keywords:
Validation protocol
Causal loop diagram
Quality of models
Verification
Socio-ecological modeling

ABSTRACT

Current modeling practices for social-ecological systems (SES) are often qualitative and use causal loop diagrams (CLDs), as these models promote an evaluation of the systems loops and variable connectivity. Our literature review demonstrated that quality assurance of these models often lacks a consistent validation procedure. Therefore, a guide to improving the validation of qualitative models is presented. The presumed utility protocol is a multi-dimensional protocol with 26 criteria, organized into four dimensions, designed to assess specific parts of the modeling process and provide recommendations for improvement. This protocol was applied to three demonstration cases, located in the Arctic Northeast Atlantic Ocean, Macaronesia, and the Tuscan archipelago. The "Specific Model Tests" dimension, which focuses on the structure of the model, revealed positive evaluations of its structure, boundaries, and capacity to be scaled up. "Guidelines and Processes", which focuses on the meaning and representativeness of the process, showed positive results regarding purpose, usefulness, presentation, and meaningfulness. "Policy Insights and Spillovers", a dimension focused on the policy recommendations, revealed a high number of "not apply", indicating that several criteria are too advanced for the status of the models tested. The "Administrative, Review, and Overview" dimension, which focused on the managerial overview, showed the models needed improvement in the documentation and replicability, while time and cost constraints were positively evaluated. The presumed utility protocol has shown to be a useful tool providing

* Corresponding author.
E-mail address: bmeirelles@azti.es (B. Oliveira).

https://doi.org/10.1016/j.crsust.2024.100273
Received 12 September 2024; Received in revised form 23 November 2024; Accepted 27 November 2024
Available online 18 December 2024

quantitative and qualitative evaluations for an intermediate evaluation of the model-building process, helping to substantiate confidence, with recommendations for improvements and applications elsewhere.

1. Introduction

The negative impacts of societal activities and pressures on ecosystems have been increasing in magnitude and complexity during the last two centuries (Rockström et al., 2009). Managing these pressures in such a way that promotes both the well-being of humankind and a healthy natural world, is one of the main challenges of the current century (Borja et al., 2024; Steffen et al., 2015; Whitmee et al., 2015). Addressing these interconnections from humankind to the natural world by understanding the causes and effects of human-nature relations is better supported through integrative research, such as the social-ecological systems (SES) analysis (Berkes and Folke, 1998; Liu et al., 2007; Ostrom, 2009). However, integrative research reveals many challenges of complex adaptive systems, including the plurality of social needs and cultures (Kinzig et al., 2013); the uncertainties regarding the interconnectivity of SESs (Walker et al., 2023); and the nonlinear behavior of its variables (Rahmandad et al., 2021). In addition, there is an urgency for science-based decision-making (Ripple et al., 2019) and in integrating different knowledge systems in management and decision-making (Funtowicz and Ravetz, 1997). In summary, ensuring the conservation of nature's structures and functions, namely the biodiversity and the flow of ecosystem services underpinning the biosphere and humankind, is crucial to the promotion of well-being, equity, and justice (Abson et al., 2014; Folke et al., 2021; Oliver et al., 2021).

The ocean has a special role in this context (Nash et al., 2017). With a crucial relevance to the health of the whole biosphere, the challenges oceans face are pervasive and multiscale, such as biodiversity loss, and the impacts of global change, among others (Borja et al., 2017, 2024). In addition, the multi-level impacts on the ocean, from the genetic structure of populations to species, habitats, and ecosystem integrity, including food webs and complex bio-physical interrelationships within the system (Borja et al., 2017), will be worsened by climate change, with consequences in terms of ecosystem services flows, vulnerabilities, and resilience (Borja et al., 2022; Oliveira et al., 2022a, 2022b). Finally, models can assist us when facing challenges of this level of complexity (Sterman, 2000) globally, particularly in marine and coastal realms.

Declaring assumptions and transcribing them into models are powerful means to assist informed decisions (Sterman, 2000). One of the main tools used to model system dynamics is the qualitative causal-loop diagrams (CLDs) (Levins, 1974; Sterman, 2000). CLDs are built to represent the 'big picture' of complex systems, by organizing their general structure, where feedback and interconnections between elements are the key components (Barbrook-Johnson and Penn, 2022a).

Due to their relatively reduced level of complexity, CLDs are broadly used in systems modeling in academia and among practitioners (Sterman, 2000). Furthermore, there is some uncertainty about how many systems thinkers submit their models for peer review and a consistent validation process. While in academia, validation is usually a requirement of the publication process, to the extent of our knowledge, this is not consistently required from practitioners, and it is often overlooked or simplified. Our literature review shows that some authors do not use a consistent process for validation of their CLDs, or usually go through a simplistic process of validation that does not enable the nuances of the modeling process to show up. As the system dynamics field has shown, the relevance of bias, the crucial importance of the boundary definition (not only geographical, but including worldviews, values, and beliefs), and administrative steps such as documentation and reproducibility are challenges for the modeling process (e.g., Sterman, 2000), and consequently are also challenges for validation. Therefore, a thorough protocol to provide confidence at intermediate stages and guidance along the modeling process is required.

In this context, this study provides a validation protocol (the presumed utility protocol) to help scientists, modelers, and practitioners increase the quality of their qualitative models, with a focus on CLDs. This protocol contributes to the current literature in the field as it embraces the most relevant criteria for validation of qualitative models, previously diffused in many sources, and complements these criteria where appropriate. In addition, our literature review will show how much the CLDs use has grown, and as this is one of the main tools for system thinking (Barbrook-Johnson and Penn, 2022b; Sterman, 2000), thereby showcasing the importance of promoting consistency of the validation practices. First, we introduce the topic and relevance of SES. Second, a short review of the nomenclature behind validation methods is presented along with the methods for construction of the protocol and its use. Third, we present the proposed protocol, and the results from the demonstration of its use in practice. Finally, by using the presumed utility protocol, we show how a nuanced perspective of the modeling process, often overlooked, can be done, which is timely and positive for the meaningfulness of the modeling process and the further use of its results.

2. Methods

2.1. Review of the use of validation practices

To verify a possible mismatch between the use of CLDs and the validation practices we consulted the Web of Science (WoS) and the Google Scholar webpages on the 9th of July of 2024 using the entry topic "causal loop diagram" in all fields and all topics. We chose these databases as they complement each other in terms of coverage (Falagas et al., 2008; Harzing and Alakangas, 2016), with the benefit of Google Scholar being free, whilst WoS is paid and was provided to the first author by FECYT (Fundación Española para la Ciencia y la Tecnología). The same research was done a second time, complemented by the entry topics "quality", "OR" "verification", "OR" "utility" "OR" "validation" in all fields and all topics to count those papers also dedicated to validation.

2.2. Notes on the semantics

Discussions regarding the quality of system dynamics models usually revolve around the terms validation and verification, which according to Lane (1995), started with the seminal work of Ackoff (1956). The foundational book Industrial Dynamics (Forrester, 1961), has a full chapter named "testing model validity", showing that validity, verification, and how to achieve them, are central to the discussions in the field. A historical perspective of system dynamics models validation (Appendix A) reveals diverse approaches for quantitative and qualitative models that, after the claims of soft systems methodologies (Checkland, 1989; Checkland and Poulter, 2020), should include different worldviews and values assumptions in model construction and validation.

Forrester (1961) claims that *"the validity (or significance) of a model should be judged by its suitability for a particular purpose."* The major issue then regards the purpose of the modeling process, which the author defines as to *"aid in the design of improved industrial and economic systems"*, consequently defining the ultimate test of validity is in the *"whether or not better [management] systems result from investigations based on model experimentation"*. The problem is this ultimate validation test might be far from the modeling process, which calls for an intermediate step of evaluation, closer to the model-building process, that helps to substantiate some confidence in it. As recommended by Churchman et al. (1957): model construction and testing should go on simultaneously.

Oreskes et al. (1994) claim that verifying or validating numerical models of natural systems is impossible. That happens for two reasons: first, these systems are open, which implies that there are variabilities in the system that necessarily were not captured by the model; second, some results, the more verisimilar they appear can be replicated by different models, and therefore it is not possible to know for sure which one represents the reality. This characteristic of models, known as indetermination (Oreskes et al., 1994) does not allow a choice between two different, but equally verisimilar, models using only as criteria the data and structure of the model; it is necessary in this case to adopt some arbitrary criteria to adopt one model or the other. Usually, these criteria are simplicity, symmetry, elegance, or even personal or political trust. But these choices per se state that it was not possible to determine which model was the truest. Verifying, thus, can only happen in closed systems when all data are known and known to be correct.

Furthermore, it is a common practice among modelers to divide data into two parts, using the first part to calibrate the model and then certifying the results of the model are coherent with that time series, and posteriorly comparing the other results with the second part of the data, from which is usually inferred that if the results and the data were congruent, the model is valid, otherwise not. This practice is misleading (Oreskes et al., 1994; Sterman, 2000) and does not ensure the validity of the model because being an open system, the congruence of data and results are occasional. In addition, it requires a numerical simulation of the results, to be compared with real data, which is not suitable for qualitative CLD models.

A common quote in the modeling field is that "all models are wrong" (Box, 1979; Sterman, 2002), usually complemented by "but some might be useful". We echo that phrase and assume it is one of the reasons why Meadows (1980) avoided the "validity or veracity" dilemma and used the term "utility" to describe the quality of system dynamics simulation, such as in Sterman (2000). Barlas (1989, 1996) also identified the limitations of the term "validity" understanding that "quality" would suit better the aim of the term. Lane (1995) also drifted from the austerity of the formal mathematical validation toward the notion of "usefulness".

Forrester (1961) still apply, that the final test of utility happens when the model unfolds into a better management system, but as it can only be determined with time, the present protocol adopts a terminology to name it as an incomplete task, as it is dedicated to evaluating something that is not fully deployed, still provisional. Echoing Meadows (1980) that the term "utility" is better for an indicator of the quality of a model, we adopt the term "presumed utility" as the indicator of the quality proposed here. The foundations below this understanding of validation are presented in Appendix A. We did not change the terms used by other authors, so validity is still present in the paper, but when we present and discuss our contribution to the topic, we will refer to the presumed utility. In summary, presumed utility is an attribute of a model, considered from the model-building process to the final use, which provides a consistent statement of confidence and quality to an ongoing management or study process.

2.3. Creation of the presumed utility protocol

To create the protocol for evaluating the presumed utility of CLDs (Table 1), we selected the contributions that were considered relevant for qualitative-like models from those available in the literature (Appendix B). The literature review started with some reference works regarding system dynamics (e.g., Sterman, 2000) and was followed by a snowball literature review. The relevance of this literature is described in Appendix A. During the review, a selection of papers that have significant contributions to validation was made. From the many indicators used for validation in the literature (Appendix B), we used a SWOT-like analysis to delimitate our sample on those indicators relevant for a qualitative model, in other words, the criterion of inclusion was that the candidate indicator could be used for assessing a qualitative modeling process (from the preparation phase of the modeling to the end use of results). The criterion of exclusion was that indicators that required quantitative information from the models to be tested (i.e., those from a numerical model) should be rejected. Finally, the list was complemented with questions to facilitate the self-assessment grading and included some tests considered relevant for us and from the literature (e.g., Meadows, 1980; Sterman, 2000), but that has not been appropriately explored so far in the literature (e.g., learning). We understand modeling broadly, as a process of discussion, learning, and formalization (in a qualitative or quantitative form) of an understanding of a system and or issue of interest, made by stakeholders during a period (see Appendix A). Therefore, validation is more than a Boolean statement of valid/invalid. To that end, these criteria (Table 1) are dedicated to learning, improving the quality, and increasing the confidence in the whole process, not only in the model, achieving in the end, the "presumed utility", the aim of the present research.

To demonstrate its usefulness, the model creation and the application of the presumed utility protocol were made in the three Demonstration Areas (DAs) of the EU-funded Horizon Europe Marine SABRES project, namely the Arctic Northeast Atlantic Ocean, Macaronesia, and the Tuscan archipelagos. The model creation process happened through a series of two-day workshops. The workshops were held online, between May and June of 2024, one for each DA. We had 10 participants from the Arctic, 4 from Tuscany, and 12 from Macaronesia, including facilitators and observers. These persons were chosen by their institutions and the project consortium as the scientist group from each DA. Most of them come from a natural sciences background, which could be a limitation of the project (discussed in section 4.2). The CLDs were built using VENSIM PLE (version 10.1) and following the guidelines from Sterman (2000) and Van den Belt (2004).

As the future analysis of these CLDs brings with it the challenges of understanding the complexity of the system, some heuristics will be used to facilitate the process. One of them is the selection of variables in the system that are perceived by the authors as more important than others. Therefore, the CLDs (Figs. 1, 2, and 3) have some variables categorized as Political, Economic, Social, Technological, Legal, and Environmental (PESTLE) (Pinnegar et al., 2021), which represent the most important variables, seen by the modelers. The content of each CLD, its PESTLE elements, scenarios, and other results will be explored elsewhere. After the model creation, the modelers (i.e., workshop participants) were invited to fill out the 26 tests (Table 1) one by one and to provide their comments. In the end, modelers attributed a grade of trust to each of the tests, using a self-report Likert scale (Jebb et al., 2021) (Appendix C) that was distributed and collected via email.

The tests (i.e., criteria) were divided into four groups or dimensions, each one dedicated to specific steps or attributes of the modeling process: (i) "Broad guidelines and processes", (ii) "Specific model tests", (iii) "Policy insights and spillovers", and (iv) "Administrative, review, and overview", organizing the type of tests contain accordingly. The first two dimensions are present in virtually any model as they represent organizational aspects, boundary discussions, and the structure produced during the modeling exercise. The third dimension is more related to the capacity of the model to create some management options and recommendations and therefore might be more suitable for the late-stage model. The last dimension focuses on the documentation, replicability, and review from a third-party perspective. In summary, dimension 1 focuses on the boundaries and meaning of the model; dimension 2 focuses on the specific structure of the CLDs; dimension 3 is dedicated to understanding the policy insights that the model has produced and finally dimension 4 encompasses the administrative, replicability, and documentation efforts.

Interpreting the results of the presumed utility criteria requires the sensibility of the evaluator of the modeling process to match the content revealed in these criteria with the issues and limitations of each model. The numerical evaluations show an overall metric (average and mode) of each criterion, but the meaning of this numerical evaluation might be

Table 1
Tests for presumed utility in qualitative models such as Causal Loop Diagrams (CLDs).

Dimension	N	Criteria	Description	Grade	Comments
Guidelines and processes	1	Purpose	The idea is to state clearly beforehand to which purpose the model will be built. Do you have a clear statement about the purpose of this model?		
	2	Usefulness	Embraces the idea of the adequacy of communication of the ideas represented in the model. Who will operate the model, the modelers or third parties? Is it available in an adequate and clear format for the user? Are they able to understand and use the model and the results? Is the model compatible with the users' capacities?		
	3	Presentation	Refers to the adequacy of the presentation of the model to the relevant audience, considering their level of scientific understanding, language, or others. Are the model and the materials used along the modeling process (such as data, tables, maps, pictures, etc.) appropriately formatted accordingly to the audience? Are the loops represented individually or in one big CLD? Is the diagram organized to reduce the number of crossed lines?		
	4	Perspectives in Boundary-adequacy	Refers to different perspectives of issues and policies. Does the modeling process support debate on different perspectives while discussing the system and its issues concerning: a) choice of model used; b) System Dynamics issue addressed; c) goals to be achieved; and d) Policies for doing so?		
	5	Norms/values in boundary adequacy	Refers to different perspectives of values and acceptability. Do the models support debate concerning and representing the behavior of the relevant actors: a) goals (are the desired states acceptable?); b) Policies (are the actions based on discrepancies between goal and actual conditions acceptable within their culture?)		
	6	Trustworthiness or Guru status of the modeling team	An affinity with the modeler can enhance positively the modeling process and the Policy Insights or Recommendations (PIoR) implementation. Is it possible to report a positive relationship or atmosphere between the stakeholders and the modeling team?		
	7	Meaningfulness of the process	Relates to the experience of stakeholders. Is it easy and fun to explore the models and search for results? How much did the relevant actors participate in the model building? How much did the relevant actors participate in the discussions regarding the model?		
Specific model tests	8	Structure-verification	By comparing the structure of the model with the [presumed] structure of the real system the model represents (considering previous questions regarding worldview and culture). Does the model represent satisfactorily the system and its issues? Are the variables stated unambiguously? Are connections representing causation instead of correlation? Are the important delays represented?		
	9	Loop Polarity	The loop polarity test compares the loops in the model with the modeler's or client's assumption about which are the relevant feedback loops in the real system. Did stakeholders identify the relevant loops? Is the polarity of the loops properly determined? Are there loops with different polarities converging in a variable of interest? Are the goals for balancing loops explicit? Are the loops named?		
	10	Boundary adequacy (as structure)	Looks for the adequacy of the aggregation level and at the same time tries to understand if the model is capturing the relevant structures of the system. Are relevant variables explicitly represented or they are aggregated (masked) with others?		
	11	Family-member	It is relative to the degree of generalization the model might have. The recommendation is that, by adjusting a few parameters, the model can reproduce a family-level behavior, instead of a case-specific behavior. Is it possible to apply this model to a similar system with minor adequations? Would it still be meaningful and useful for the creation of policy insight or recommendations?		
	12	Extreme conditions	Despite this being relative to the numerical model, it is brought here because the structure of the model can allow some inferences for plausible extreme combinations of state variables. Would the model presumably behave properly if variables assume extreme conditions? Is it possible to infer this from the present model?		
Policy insights and spillovers	13	Insight generation capacity	Whether a model does lead to any PIoR. Did the model lead to any policy insight or recommendation?		
	14	Relevance and Fertility of PIoR	Whether the policy insight or recommendation is innovative and important. Does the policy insight or recommendation represent an innovation to managing the system? Is the PIoR relevant?		
	15	Congruence of PIoR with culture	This test verifies the social implementability of any policy insight or recommendation. The point is that makes no sense to propose actions/policies that involve actions considered unacceptable or unbearable for a potential observer. Is the PIoR acceptable to all involved in the modeling process?		
	16	Boundary adequacy (as policy)	Concerns testing how the change in the boundaries of the model would affect the policy recommendations created by the simulation. In addition, the same policy can be tested for its adequacy if implemented outside the original boundaries set in the model. Would the PIoR require change if applied to a different location? How would the PIoR behave if applied to a larger system?		
	17	Learning	Do participants state that they learned about the system, other stakeholders (the community), or the policy-making process during this modeling process? Are they satisfied with that? If they want to learn more, did they receive support on how to do that?		
	18	Engagement	Did stakeholders engage in any group/action related to the issues dealt with in the modeling, during/after the modeling exercise?		
Administrative, review, and overview	19	Ease of Enrichment	Concerns about the ability of any model to be updated with new data, or used to test the effects of new policies. How easily can this model be complemented by new information or complementary issues in the system?		
	20	Time & Cost of the Intervention	Should be measured against a target and inform the level of satisfaction with the results against the target investment. Was the modeling process concluded within the		

(*continued on next page*)

Table 1 (*continued*)

Dimension	N	Criteria	Description	Grade	Comments
			expectations of time and costs? Are there recommendations to improve the efficiency of the modeling exercise for the next team or exercise?		
	21	Documentation	Refers to the adequacy of the process of making every step in the modeling process replicable by taking a formal process or writing assumptions, discussions, updates, or a change in previous steps regarding the modeling process. Is the model satisfactorily documented?		
	22	Replicability	Refers to the capacity of a third party to reproduce the model based on documentation. Are you sure that independent third parties can reproduce the model and all the results only using the written documentation?		
	23	Audit or cross-validation	Measure how adequately a model study is conducted concerning established standards, practices, guidelines, or experience. Preferably done by someone not involved in the modeling process. Consider differences in culture before applying this. Does the model and PIoR make sense? Are they contradicting any physical law or rigorous social norms that turn the model/PIoR invalid? Are they contradictory with experience beyond an acceptable level?		
	24	Higher-level Model review	A higher management level test of the model's appropriateness to the systems definition and study objectives, adequacy of underlying assumptions, adherence to standards, modeling methodology used, model representation quality, structure, completeness, consistency, and documentation. Preferably answered by someone at a higher level than the modeler team. Does the model fulfill the expectations of the proposed modeling exercise?		
	25	Walkthroughs	Represent group exercises dedicated to testing the overall documentation for any errors. Does not test performance. Preferably answered by a small group different than the modeler team. Does the model seem correct? Does the documentation allow the reproducibility of the model? Are the main issues represented satisfactorily? Does the PIoR make sense, if applied?		
	26	System-improvement	Considers whether the behavior of a system improved after the implementation of the policies tested in silico. It is recommended to verify this with some indicators of the desired state of the system. Is it possible to connect some changes in the system to the modeling exercise? Are these changes congruent with the desired state modeled?		

complemented by the comments for each criterion provided by the participants. In the present case, we analyzed the metrics, complemented them with an evaluation of the comments made to each criterion, and finally created a “recommendation for improvement” that reflects the main improvement required by the modelers.

2.4. *Context about the sites (demonstration areas)*

This protocol was tested in three diverse SESs in Europe. These areas were chosen by the authors of the research with the rationale they represent the relatively simple SES in Europe (in terms of the number of human activities, pressures, and interactions) and were considered suitable to be explored scientifically under the SES lens. The first site, Macaronesia, includes the archipelagos of the Azores, Madeira, and Canary Islands, hotspots for tourism and biodiversity (Parretti et al., 2020*;* Sambolino et al., 2022). The challenge here is to balance marine conservation with tourism and maritime recreation, by possibly promoting the creation of an ecological corridor connecting these archipelagos. Current conflicts of economic use and conservation are also reflected in the economic necessities regarding supporting marine protected areas. The model embraces these aspects and combines them with the presence of local and migratory species and the spill-over effects of marine conservation.

The second site, the Arctic Northeast Atlantic, is represented by the Exclusive Economic Zone EEZs and shared regulatory areas formed by East Greenland, Iceland, and the Faroe Islands (Rigét et al., 2019). The marine ecosystem plays a significant role in these countries ́economies due to the high level of commercial fishing activities, Distribution of commercially important and widely distributed pelagic stocks is often dynamic. Changes in stock presence within a country's EEZ have a direct impact on the fishing industry. Therefore, the model was dedicated to delineating a set of social and environmental challenges related to pelagic fisheries, the challenges of international political consensus regarding fish quota sharing agreements, and the complications that climate change impacts can bring to the system (Predragovic et al., 2024), among others.

The third site is the Tuscan archipelago, a large marine park in Europe and a UNESCO site (Renzi et al., 2010). The Tuscan archipelago is composed of seven islands managed differently according to society's needs. Tourism is the most important issue here (with about 500 thousand visitors per year), with which some ecological impacts are known to be associated, such as nutrient discharge, seabed disturbance due to anchoring, pollution, and invasive species. The model constructed here focused on the conservation/restoration of the seagrass meadows (*Posidonia oceanica* (L.) Delile) and associated ecosystem services, such as coastal protection, carbon storage, habitat provisioning, and others.

3. Results

3.1. *Review of the use of validation practices*

Our consultation on the Web of Science (WoS), returned 995 articles, in more than 100 fields. Environmental sciences was the most relevant, with many of the articles dedicated to the marine/coastal environment (e.g., Kapsalis, 2022; Mousavi et al., 2023; Oliveira et al., 2022b; Payo et al., 2016; Sanò et al., 2014). The use of CLDs as a research topic increased almost 30-fold from 2001 to 2005 (16 papers) to 2020–2024 (448 papers). A Google Scholar search with the same entry revealed a greater number of references (195,000) with a similar pattern, revealing a 6.5-fold increase in publications (from 995 to 6450 references) and a 20-fold higher absolute number when compared to WoS. Nonetheless, the WoS literature database search revealed that less than 13 % (129 articles) used words related to CLD's quality, utility, or validation in their topic, which called attention to an asymmetry between the use of these models and ideas of quality and confidence in their results. We speculate that this asymmetry is not due to inconsistency in their methods but to the rarity of a consistent practice of assuring the validity of the CLDs, using a systematic approach. The numbers suggest that the concerns regarding CLDs ́ quality assessments have not evolved at the same pace as their use, which makes the present study, and the protocol it presents, timely.

3.2. Causal loop diagrams for each demonstration area

Whilst an in-depth analysis of these models will be done elsewhere, a short description is provided here to make clearer the content of each model validated.

3.2.1. CLD model for Macaronesia

The Macaronesia region model integrates two countries (Portugal and Spain) and three archipelagos (Azores, Madeira, and the Canary Islands). It contemplates the two most pressing issues identified by the DA for the region: tourism and its impacts, and the creation of a protected ecological corridor connecting the three archipelagos (Fig. 1).

In this CLD, human pressures negatively affect the habitat quality (main environmental variable), which in turn can affect others, such as marine birds, mammals and turtles, migratory fish, tourism, and welfare. Marine birds, mammals, and turtles are relevant for tourism activity, but also to the cultural heritage of the archipelagos (along with fisheries). Tourism and fisheries are relevant for economic development (main economic variable) and welfare (main social variable). Welfare is related to consumption, and depending on the human population density, would influence the pressures on the environment. Consumption is influencing marine traffic, as most of the goods reach these archipelagos by boat. Marine traffic can have environmental influences such as collision with mammals, noise and other forms of pollution, facilitating invasive species colonization, etc., also having a negative influence on the consensus for the creation of the protected corridor (the main political variable), as these vessels could have increased their traveling costs under the possibility of the creation of an ecological corridor with restrictive access for these vessels. Geopolitics is relevant for its influence in this consensus, but also on the level of protection (the main legal variable) of the protected areas in the region, which would include the corridor. Geopolitics, the level of protection, and the technological advances from nautical and navigation technology variables are perceived as exogenous variables, i.e., they are not influenced or controlled by other variables within the model. The creation of this protected corridor is seen as positively influencing ecological connectivity, which would, in turn, promote the abundance of marine birds, mammals, and turtles. Nautical and navigation technology (the main technological variable) was seen as facilitating the political consensus for the corridor and preventing and reducing the negative impact of collisions, pollution, and invasive species.

3.2.2. CLD model for the Arctic Northeast Atlantic

The Arctic DA produced a CLD that integrated the pelagic fisheries issues in the Northeast Atlantic Ocean which is a shared area for the three countries present in this study (Fig. 2).

The model represents an agreement on fish quota allocation as the main political variable, and after the workshops' discussions, this is one of the main topics this DA wants to explore in future analysis and further development of the study. This agreement is influenced by political consensus, influenced by high-level government decision-making (the main legal variable), other countries' goals, the discrepancies between the expected and real decisions taken in the past regarding allocation, and the power of the fishery sector. The agreement is also influenced by the variability of fish distribution, and the predictability of productivity, environmental variables which are highly uncertain and influenced by climate variability (main environmental variable). As a rational agreement on quota sharing is provided, it influences positively the stock of fish due to the rational and efficient exploitation of the resource, reducing the fishing effort, which is related to the landings (fish on land). Landings (the main economic variable) can negatively influence the stock of fish in the long term (note the delay mark $\neq$) if overfishing occurs, but also directly support exports and local development (which are both the most important social variables). Landings also influence jobs, and exports, which along with profit, are the main drivers of taxes and then enhance the national economy. Other jobs are also relevant for the economy, and they are related to fisheries gear companies, which in turn promote the fishery technology sector (the main technological variable), with feedback in job creation.

3.2.3. CLD model for Tuscan Archipelago

The Tuscan Archipelago DA produced a model addressing tourism and other pressures on the seagrass *Posidonia oceanica* (Fig. 3).

The story in the model starts with the tourism activity. Tourism is negatively influenced by limitations on human activities within Marine Protected Areas (the main legal variable). Tourism is central in the

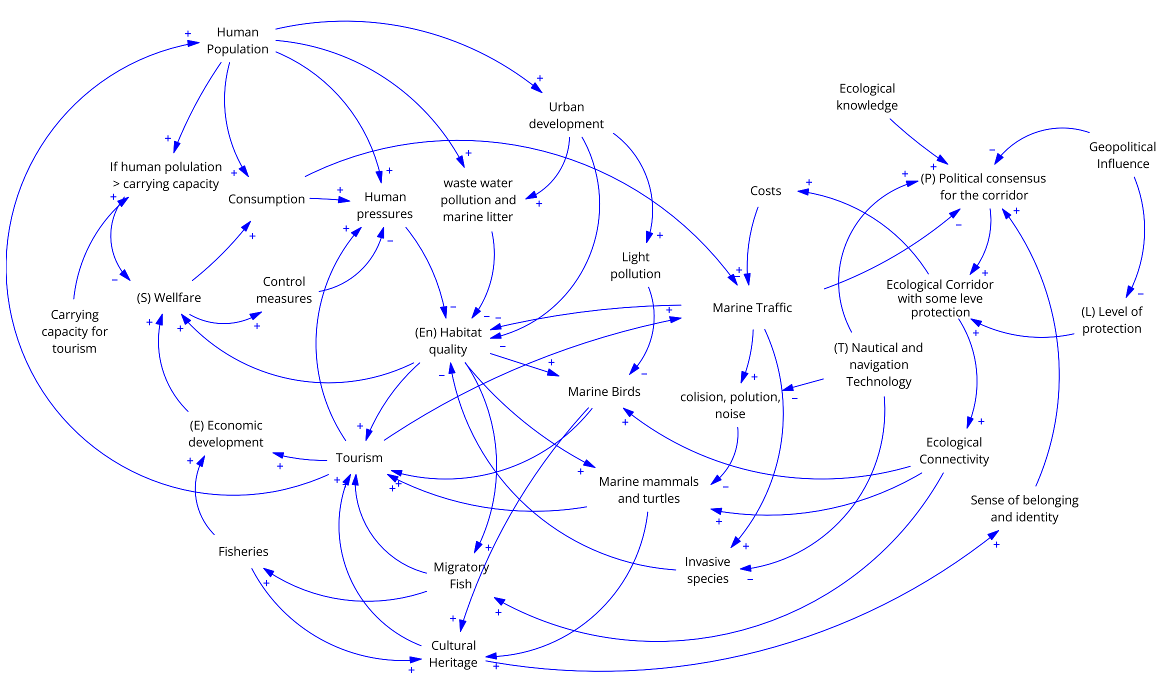

Fig. 1. Causal Loop Diagram for the Macaronesia Demonstration Area. +: positive effect; -: negative effect; (P): Political, (E): Economic, (S): Social, (T): Technological, (L): Legal, and (En): Environmental

model, as it influences plastic pollution, boat traffic, and nutrient discharge, but also the local economy, diving centers' income, boat renters' income, and the generation of inflation (represented by the variable prices). Boat traffic leads to anchoring in the coastal zones, which is one of the causes of the regression of seagrass meadows (the main environmental variable) and influences negatively the efforts of restoration of the seagrass. Seagrass meadows influence positively carbon sequestration (mitigating, with delays, climate change, and extreme events), and provide other benefits such as erosion control, public services, fauna and flora biodiversity, health, control of invasive species, and promote nutrient cycling. Health is important to the well-being of locals (the main social variable) which is also influenced by public services, prices, jobs (the main economic variable), and tourists. Jobs influence taxes (the main economic variable), which play a crucial role in funding public services. An app organizing tourism activities (the main technological variable) would be relevant to reducing the anchoring in coastal zones and its negative effects on the seagrass meadows. Health is also negatively influenced by pathogens that come from nutrient discharges in the water and from the impacts of climate change. The nutrient discharges that come from tourism activity and boat traffic, are controlled by the sewage treatment plants, regulated by land-distance norms for discharging, and influence the water quality which inspired the blue-flag program (a quality label for beaches and other coastal areas that can promote tourism activity) for safe tourism. Nutrient discharge and anchoring can reduce the chances of success for *P. oceanica* restoration activities, and therefore all the benefits these meadows produce would be diminished.

3.3. Presumed utility protocol results

Of the 26 workshop participants, 16 completed the tests in the utility protocol. When some criterion was blank, it was considered as "not apply" (<3 %). Answers provided more than 400 numerical evaluations (Table 2 and Appendix C) and around 300 comments. The number of responses was considered satisfactory for an appreciation (its validation) of the proposed method. We encourage future users to reflect on the meaning and relevance of the comments to evaluate which number of answers is suitable for each application. All comments are integrally available for consultation (Appendix D).

The overview balance of the evaluations considered the modeling process positive with a low number of evaluations (7 %) being of "very dissatisfied" and "moderately dissatisfied", against an overall satisfaction (37 %) composed of "moderately satisfied" (27 %) and "very satisfied" (10 %). Answers that considered the criterion "do not apply" were (32 %), and 24 % were neutral ("neither satisfied nor dissatisfied").

The four dimensions categories of indicators ("Guidelines and Processes", "Specific Model Tests", "Policy Insights and Spillovers", and "Administrative, Review, and Overview") presented different levels of satisfaction that point to different aspects and stages of the modeling process (Fig. 4). The first dimension had many more positive evaluations (50 %) when compared to negative ones (6 %). The average results varied from 3.3 (boundary adequacy) to 4.1 (Meaningfulness of the process). The second dimension had more positive results (55 %) than the sum of all other categories of answers. The average results varied from 3 (extreme conditions) to 4.1 (boundary adequacy). In the third dimension, the average result varied from 2.9 (insight generation capacity) to 3.8 (learning). Here the number of "not apply" was larger than other dimensions (48 %). Lastly, in the fourth dimension, the average values vary from 2.5 (Replicability) to 3.9 (time and cost of intervention). In this dimension the NA was large (43 %) but smaller than the previous one, followed by the positive evaluations (26 %).

In general, the number of "not apply" (NA) starts with a high level in

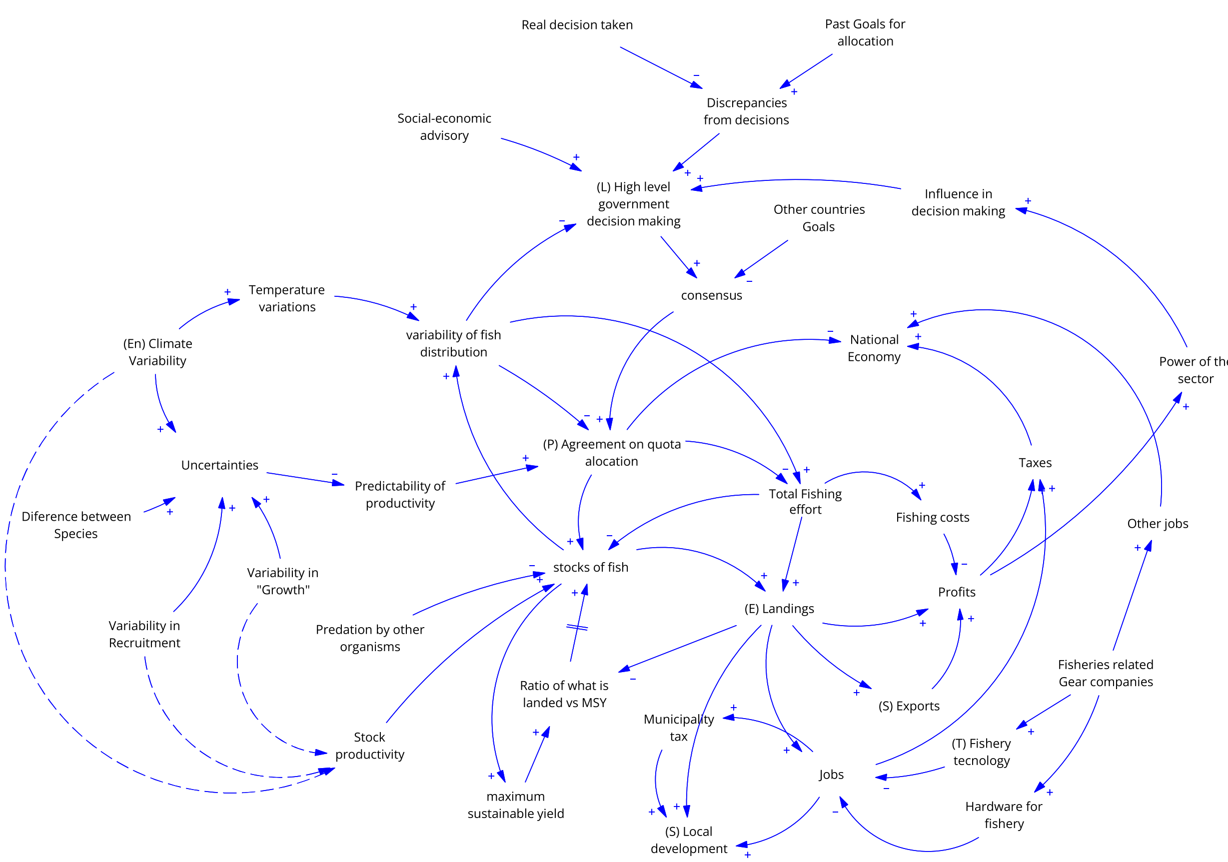

Fig. 2. Causal Loop Diagram for the Arctic Northeast Atlantic Demonstration Area. The symbol ≠ is a delay mark signaling the effects of one element in the next one happens with significant delay. +: positive effect; -: negative effect; MSY: Maximum Sustainable Yield; (P): Political, (E): Economic, (S): Social, (T): Technological, (L): Legal, and (En): Environmental

dimensions 1 and 2 (19 %) that come mainly from criterion 6 (status of the modeling team), 9 (Loop polarity), and 12 (extreme conditions) which were not explored in depth in the present modeling process. For the last two dimensions the number of "not apply" increases possibly indicating that some criteria are premature to the status of the modeling exercise used here. The numerical evaluation of each criterion (Table 2) is an average from those obtained by the protocol. From the comments (Appendix D) we elaborate on the main recommendation for each criterion.

4. Discussion

4.1. General topics

The necessity of coordinating efforts in creating models that are meaningful and somehow endeavor a validation process is present in many fields outside the system dynamics or causal loop diagrams practice and forms an important standardization of practices. An example of these protocols, such as the ODD (Overview, Design concepts, and Details) (Grimm et al., 2006, 2020), was focused on promoting the replicability of Agent-Based models (ABM), considered a relevant scientific practice, especially in quantitative models. We build on this idea of standardization of validation practices to propose the presumed utility protocol.

The presumed utility protocol was designed to be epistemically broad, including the descriptions and analyses of model replicability (e.g., Criterion 21, Documentation; Criterion 22, Replicability) as the ODD, but not as the most relevant features of the process. As the main objective is to promote learning, it builds on facilitating that the users have elements to reflect upon the process they are taking part in. As a reflective process, modeling becomes a process of learning (Forrester and Senge, 1980; Sterman, 2000), and therefore the outcomes of this process are not only the model per se but the cognitive transformation it invites on those involved in it.

The mode was also presented as it can be an easy proxy to test how informative is the average in representing the evaluations. In our case, the low value of mode in indicators 23–26 makes sense regarding the model status. In addition, on criteria 9 (loop polarity), 12 (extreme conditions), 14, and 15 (dependents on policy insights), and 18 (dependent on public engagement) the mode is congruent with the premature status regarding these topics, and that some practices that will be done in future analysis of the model (i.e., loop analysis).

4.2. Multidimensional utility protocol

The first two dimensions had more positive results compared to the last two. The first reflects their satisfaction with the meaningfulness of the process and the boundaries of the model, which were well explored during the workshops. The second dimension refers to the structure of the model, which they are also satisfied with, where the main exception was the loop polarity, intentionally left for further analysis of the modeling results (not in the scope of this paper). The third dimension might require a higher status (i.e., a later stage in development) of the management process where policy recommendations were already produced and applied. Finally, the administrative dimension will be expected to show higher satisfaction from those models that are in a stage of maturity where even the application of its policy recommendations was already made, and the documentation was made available after a structured reproducibility assurance process. Both dimensions 3 and 4 have criteria that were revealed to be a bit premature for the case

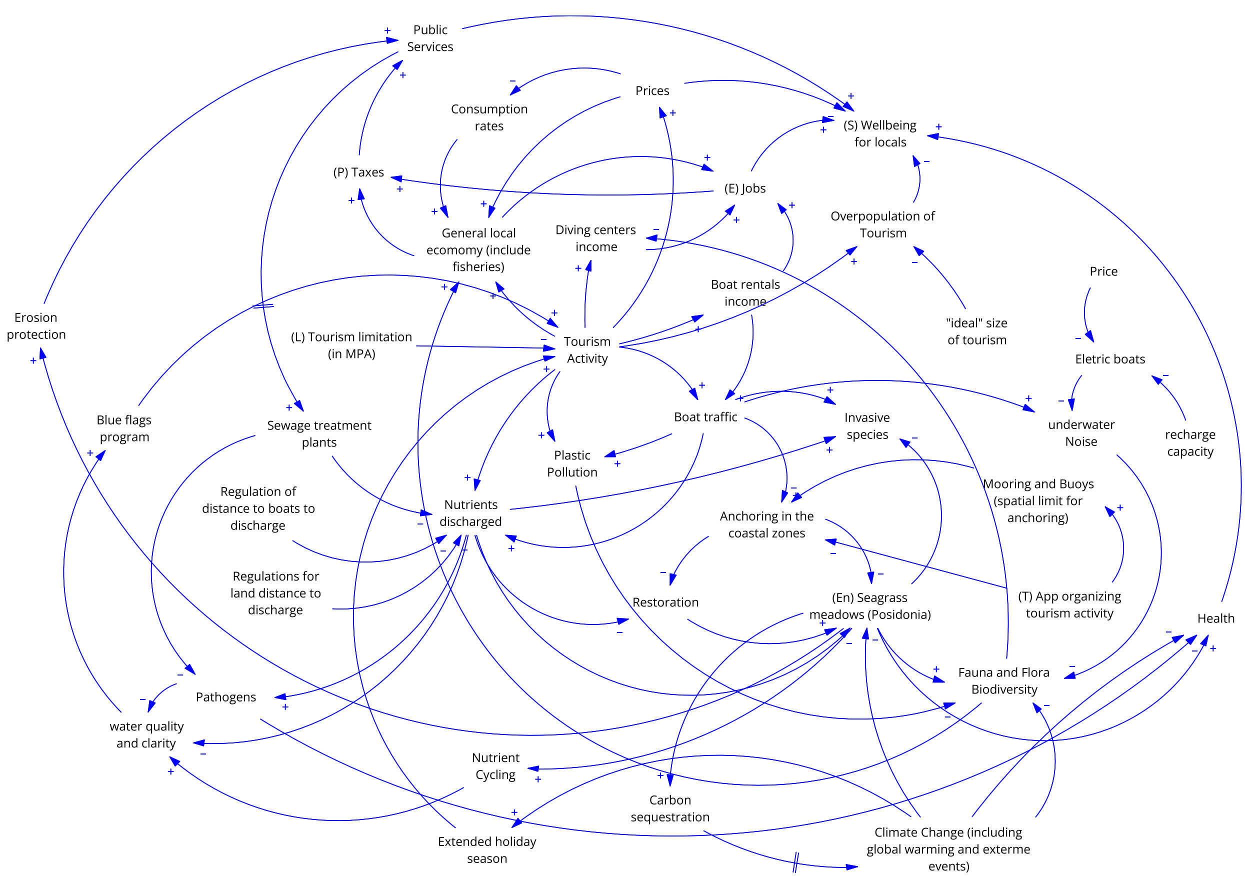

Fig. 3. Causal Loop Diagram for the Tuscan Archipelago Demonstration Area. The symbol ≠ is a delay mark signaling the effects of one element in the next one happens with significant delay. +: positive effect; -: negative effect; MPA: Marine Protected Area; (P): Political, (E): Economic, (S): Social, (T): Technological, (L): Legal, and (En): Environmental

Table 2
Results from quantitative evaluations on each presumed utility's criteria. Dim: dimension, N: number of the criteria; Avg: average value; M: maximum value, CLD: Causal Loop Diagram; PIoR: Policy Insights or Recommendations.

Dim.	N	Criteria	Avg	M	Recommendation
Guidelines and processes	1	Purpose	3.9	4	More time discussing the purpose of the model.
	2	Usefulness	3.6	4	Introduce previously the problems, definitions, delimitations, and how to navigate a CLD.
	3	Presentation	3.5	3	Simplify the presentation to individual loops and associate it with maps, graphs, pictures, or others.
	4	Perspectives in Boundary-adequacy	3.3	3	Include more people with different perspectives in the modeling exercise.
	5	Norms/values in boundary adequacy	3.8	4	Include people with different perspectives in the exercise.
	6	Trustworthiness or Guru status of the modeling team	4.0	0	Increase the connections with stakeholders to promote a better understanding.
	7	Meaningfulness of the process	4.1	0	Show the model to stakeholders and other relevant actors.
Specific model tests	8	Structure-verification	3.8	4	Clarify the difference between causation and correlation and why using causation for structuring the model.
	9	Loop Polarity	3.6	0	Create a group understanding of the loops.
	10	Boundary adequacy (as structure)	4.1	4	Discuss aggregation and disaggregation with the group.
	11	Family-member	3.9	4	Build the model in groups to increase its level of generality (promoting its usefulness to a general case).
	12	Extreme conditions	3.0	0	At future users' discretion, remove this criterion to reduce confusion.
Policy insights and spillovers	13	Insight generation capacity	2.9	3	Ask the modelers what the main messages the model is passing are and what it implies for policymaking.
	14	Relevance and Fertility of PIoR	3.1	0	If any policy recommendation was provided, ask the modelers to evaluate its relevance.
	15	Congruence of PIoR with culture	3.6	0	Start a consultation process exposing the recommendations (considering the variations of people's views, and culture).
	16	Boundary adequacy (as policy)	3.0	3	Explore the spectrum of policy insights concerning the user's culture and with broader goals of equality, and justice.
	17	Learning	3.8	0	Better declare who are the foci of this indicator and help the modelers provide feedback on what they learned.
	18	Engagement	3.5	0	Extrapolations of the knowledge in the modeling process can increase the reach of the policy recommendations.
Administrative, review, and overview	19	Ease of Enrichment	3.4	4	Use tools with low technological requirements.
	20	Time & Cost of the Intervention	3.9	4	Use more interactive tools during workshops.
	21	Documentation	3.4	3	Recording the workshops might help with documentation. Caution as it can be coercive to people with different views.
	22	Replicability	2.6	3	Having a manual for documenting the experience or a facilitator writing every step in the exercise.
	23	Audit or cross-validation	3.4	0	Guide the reviewer to be strict about physical laws but flexible about social norms, culture, or views.
	24	Higher-level Model review	3.6	0	Offer this protocol to the higher-level reviewer or coordinator of the project.
	25	Walkthroughs	4.0	0	Open to any people who want to provide feedback on the model to make a walkthrough.
	26	System-improvement	0.0	0	Open to the possibility of having some practical results, even in early-stage models.

study the present validation was dedicated. This has not been shown as a problem for the validation protocol, as its structure showed versatility in capturing this information.

The spirit behind the **first dimension** ("Guidelines and process") was to first make clear to the modelers that they understood the purpose of the exercise and that it is meaningful, as it is an old premise in the field (e.g., Meadows, 1980), but sometimes overlooked due to other pressing tasks in the modeling. Secondly, it opens to plural worldviews, values, and perspectives about the system, which has a long history in the field as exemplified by soft system methodologies (Checkland, 1989; Checkland and Poulter, 2020), and are currently gaining attention in more recent research such as in the social-ecological systems (Bennett et al., 2021; Berthet et al., 2022; Oliveira et al., 2024).

The main recommendations in the first dimension (Table 2) reiterate the relevance of a clear purpose (echoing Schwaninger and Groesser, 2020; Sterman, 2000), and initial "training" of the users about navigating CLDs and using them for decision-making. Simplifying the presentation was also pointed out as relevant here, and resonant with the literature (e.g., Balci, 1994; Sterman, 2000). An important aspect of our results pointed to the necessity of broadening participation in the modeling exercise (Criteria 4–7). These criteria were important firstly because the modelers identified the lack of variety within their group (their own bias). As the groups in each area are mostly formed by natural scientists somehow connected to the marine topic, this identification is relevant and coherent. As extensively described and explored in models (Checkland, 1989; Checkland and Poulter, 2020), plurality should be explored as they provide complementary views of the system. The result in this dimension is positive as raising awareness of the limitations of the group toward the whole decision-making process, diminishing what Ravetz (1993) called ignorance of ignorance (i.e., the leap one siloed group of scientists makes by assuming their rationality is universal).

The recognition by the participants of a necessary plurality is also related to concepts of plural peer community (Funtowicz and Ravetz, 1997) and is timely for the marine social-ecological systems governance. As highlighted by current research in the field (e.g., Schadeberg et al., 2024), the legitimacy provided by plural participation in the governance process is crucial to complement scientific knowledge to acquire social licensing for using marine areas.

The **second dimension** ("Specific model tests") converges with the most critical recommendations in the modeling field (Barlas, 1996; Forrester and Senge, 1980; Sterman, 2000) regarding the structure of the model and the polarity of loops. Additionally, it questions the possibility of this structure to represent a generalized property, extrapolating the ideas revealed by each case study.

Structure-verification is one of the main indicators that the model under scrutiny makes sense, as it implies the users double-checked its structure and it is both intelligible and reflects the issues they perceive relevant in the system (e.g., Barlas, 1989, 1996; Cassidy et al., 2021; Crielaard et al., 2022; Forrester and Senge, 1980; Lane, 1995; Schwaninger and Groesser, 2020; Sterman, 2000). Additionally, the questions regarding the difference between causality and correlation and aggregation/disaggregation in the structure provided recommendations

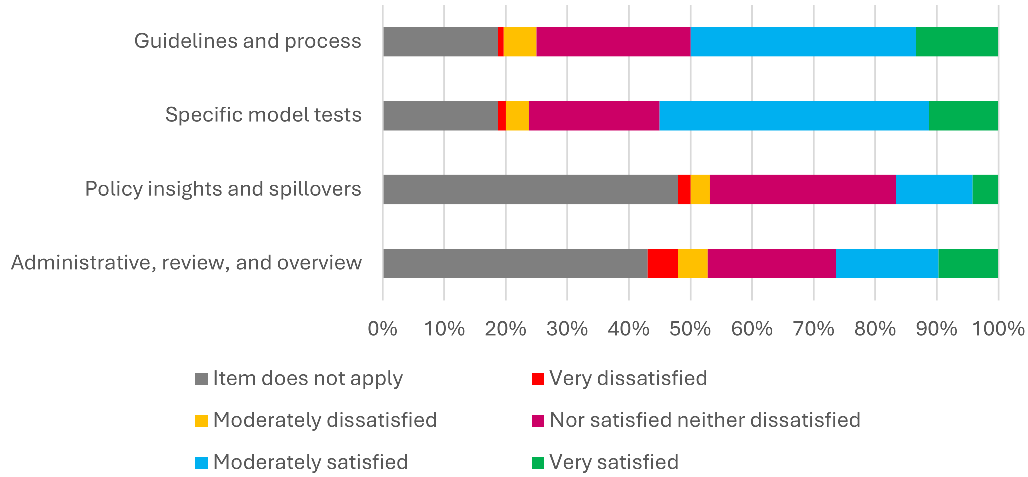

Fig. 4. Distribution of evaluations per dimension.

about the necessity of better discussions on these topics with the group. Models using causation are different from those using correlation (Overmars et al., 2007), and using correlation for the creation of CLDs is misleading and should be avoided (Sterman, 2000). As the loop polarity will be explored in the applications of this model, it was superficially explored here. Extreme conditions were revealed to be confusing as people were expecting a numerical model to use it, and therefore it can be removed from the protocol in future uses. The capacity of the present models to represent a general case (Forrester and Senge, 1980; Schwaninger and Groesser, 2020; Sterman, 2000) was considered positive with some adjustments and here the comments pointed that more participation would increase its capacity for generalization.

The **third dimension** ("Policy insights and spillovers") evaluates the outputs that came from the modeling exercise and its adequacy to the boundaries of the system under scrutiny (i.e., the culture of the people in the system it is dedicated to investigating). In addition, it iterates on the meaningfulness of the process by checking for learning and the possibility that the people in the system are promoting some sort of collective actions (Ostrom, 2000), or social learning (Reed et al., 2010) promoted by the modeling process.

The recommendations on this dimension were limited due to the premature status of the model to generate insights, yet two-thirds of the DAs did produce insights (Appendix E, question B). This is positive as these insights represent one of the main final uses of the model (Forrester and Senge, 1980; Lane, 1995). The congruence of these insights, though, can be tested in the future with other representatives of these regions, as recommended (Table 2). Comments about learning indicate the users of the protocol need clarification about the foci of this question (some of them considered the people in the system and not the modelers as the foci), but overall indicators revealed people learned during the process, which is important taken modeling as a learning practice (Sterman, 2000).

The **fourth dimension** ("Administrative, review, and overview") is dedicated to documentation, replicability, cost evaluations, and other relevant elements to the process (Schwaninger and Groesser, 2020; Sterman, 2000). It also embraces a criterion for system improvement, which can be meaningful to advanced modeling processes.

The recommendations pointed to the use of more interactive tools during the workshops, and for the necessary simplicity of the modeling tool to certify the model can be reproduced and enriched with minimum costs and effort, as recommended by Lane (1995). Enrichment is important as a learning step (Sterman, 2000) as it brings different knowledge that should be embraced by the model. Regarding documentation and replicability, the recommendations focused on the possibility of recording the workshops, but we caution that this must be done carefully due to potentially sensitive information and always must consider data protection plans. Recommendations of auditing, third-party walkthrough, and other reviews were taken positively by the groups but should consider the safeguards of sensitive information and respect different worldviews (Checkland and Poulter, 2020). Finally, the status of the model was considered premature to look for improvements in the real world. It was acknowledged the a low probability of having such results in early-stage models (Sterman, 2000), but as the performance in this indicator can improve in a later-stage model, this criterion should be kept in future uses of the protocol. Lastly, a one-by-one discussion of all criteria is available in the supplementary material (Appendix F).

4.3. Limits and caveats of the methodology

The presumed utility of models starts from a place of curiosity and learning, it is not an authoritative over-reaching statement and cannot ensure the future success of modeling and management. The first limit we point out here is that the content of the present study refers to the realm of qualitative models of system dynamics. Models that are qualitative and come from different backgrounds (not necessarily CLDs, but also the theory of change models, fuzzy cognitive maps, or others) might profit from this protocol as well but its usefulness for quantitative models is limited. For methods in statistical validation, consult the literature cited (e.g., Barlas, 1996; Schwaninger and Groesser, 2020). For a criticism of qualitative models visit Richardson (1986, 1997). Regarding archetypes critique see Forrester (1994), Lane and Smart (1994), and Richmond (1994). Criticism about the Soft Systems Methodologies (SSM) can be found in Forrester (1994), which claims SSM to be a part of the system dynamics process. This reinforces the statement that hard systems and soft systems are compatible (Reisman and Oral, 2005). Nonetheless, it is recommended to those interested in broadening the worldviews adopted by the models, namely, to embrace conflicting worldviews, to go through the SSM literature due to the specific role these perspectives have in a modeling approach (e.g., Checkland and Poulter, 2020).

Another limitation of the present research is the state of the maturity of the models. As time constraints limited our capacity to refine these models, what we present is the result of a first iteration of a modeling exercise, and the validation process after this one round of modeling. With more iterations, the model would increase its quality, and the results declared in the presumed utility protocol would evolve accordingly. Finally, as the group of stakeholders that created the model and used the present protocol to validate both process and results come mostly from the natural sciences, there are epistemological limitations and some biases that are not under the control of the present study. The stakeholders are aware of this limitation and discussed it in section 4.2.

5. Conclusions

This paper provides a tool for evaluating the quality of CLDs. We applied the presumed utility protocol in three demonstration areas of the EU-funded Marine SABRES project, namely the Arctic Northeast Atlantic Ocean, Macaronesia, and the Tuscan archipelago. Most positive evaluations were shown for the first two dimensions "Specific Model Tests" and "Guidelines and Processes", revealing the criteria in these dimensions were most suitable to the status of the model tested. The third "Policy Insights and Spillovers" and the fourth dimension "Administrative, Review, and Overview" revealed a high number of "not apply" indicating that several criteria are too advanced for the status of the tested models.

All criteria provided invaluable insights into the modeling process toward reaching the final goal: improving the quality of the system under scrutiny. Presumed utility protocol has shown to be a useful tool to provide quantitative and qualitative evaluations for an intermediate step of the model-building process, helping to substantiate some confidence in it, with recommendations for improvement. As this protocol is broad and includes the most relevant aspects found in the literature related to qualitative models, it can potentially be useful to other types of qualitative models beyond causal loop diagrams and guide the modeling process toward clearer and better models.

Funding

This research has been made in the framework of the Horizon Europe project Marine SABRES ("Marine Systems Approaches for Biodiversity Resilience and Ecosystem Sustainability"), funded by the European Union under Grant Agreement n° 101058956.

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

The authors would like to thanks to Dr. Emma Verling and Dr. Zacharoula Kyriazi (University College Cork – National University of Ireland) for their collaboration as observers during the workshops. The authors would also like to thank Dr. Lisandro Benedetti-Cecchi (University of Pisa, Italy) for the important comments and reviews made in the manuscript. This is contribution n° 1251 from AZTI, Marine Research, Basque Research and Technology Alliance (BRTA).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crsust.2024.100273.

Data availability

Data will be made available on request.

References

Abson, D.J., von Wehrden, H., Baumgärtner, S., Fischer, J., Hanspach, J., Härdtle, W., Heinrichs, H., Klein, A.M., Lang, D.J., Martens, P., Walmsley, D., 2014. Ecosystem services as a boundary object for sustainability. Ecol. Econ. 103, 29–37. https://doi.org/10.1016/j.ecolecon.2014.04.012.

Ackoff, R.L., 1956. The development of operations research as a science. Oper. Res. 4 (3), 265–295.

Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. Ann. Oper. Res. 53 (1), 121–173. https://doi.org/10.1007/BF02136828.

Barbrook-Johnson, P., Penn, A.S., 2022a. Causal Loop Diagrams. In: Barbrook-Johnson, En P., Penn, A.S. (Eds.), Systems Mapping: How to Build and Use Causal Models of Systems. Springer International Publishing, pp. 47–59. https://doi.org/10.1007/978-3-031-01919-7_4.

Barbrook-Johnson, P., Penn, A.S., 2022b. Systems Mapping: How to Build and Use Causal MODELS of systems. Springer Nature. https://library.oapen.org/handle/20.500.12657/57376.

Barlas, Y., 1989. Multiple tests for validation of system dynamics type of simulation models. Eur. J. Oper. Res. 42 (1), 59–87. https://doi.org/10.1016/0377-2217(89)90059-3.

Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. Syst. Dyn. Rev. 12 (3), 183–210. https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4.

Bennett, E.M., Morrison, P., Holzer, J.M., Winkler, K.J., Fraser, E.D.G., Green, S.J., Robinson, B.E., Sherren, K., Botzas-Coluni, J., Palen, W., 2021. Facing the challenges of using place-based social-ecological research to support ecosystem service governance at multiple scales. Ecosyst. People 17 (1), 574–589. https://doi.org/10.1080/26395916.2021.1995046.

Berkes, F., Folke, C. (Eds.), 1998. Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience. Cambridge Univ. Press.

Berthet, E.T., Bretagnolle, V., Gaba, S., 2022. Place-based social-ecological research is crucial for designing collective management of ecosystem services. Ecosyst. Serv. 55, 101426. https://doi.org/10.1016/j.ecoser.2022.101426.

Borja, A., Elliott, M., Uyarra, M.C., Carstensen, J., Mea, M., 2017. Bridging the gap between policy and science in assessing the health status of marine ecosystems. In: En Frontiers in Marine Science, Vol. 4. Frontiers Media SA, p. 32.

Borja, A., Pouso, S., Galparsoro, I., Manca, E., Vasquez, M., Lu, W., Yang, L., Uriarte, A., 2022. Applying the China's marine resource-environment carrying capacity and spatial development suitability approach to the Bay of Biscay (North-East Atlantic). Front. Mar. Sci. 9 https://www.frontiersin.org/articles/10.3389/fmars.2022.972448.

Borja, A., Elliott, M., Teixeira, H., Stelzenmüller, V., Katsanevakis, S., Coll, M., Galparsoro, I., Fraschetti, S., Papadopoulou, N., Lynam, C., Berg, T., Andersen, J.H., Carstensen, J., Leal, M.C., Uyarra, M.C., 2024. Addressing the cumulative impacts of multiple human pressures in marine systems, for the sustainable use of the seas. Front. Ocean Sustain. 1, 1308125. https://doi.org/10.3389/focsu.2023.1308125.

Box, G.E., 1979. All models are wrong, but some are useful. Robust. Stat. 202 (1979), 549.

Cassidy, R., Tomoaia-Cotisel, A., Semwanga, A.R., Binyaruka, P., Chalabi, Z., Blanchet, K., Singh, N.S., Maiba, J., Borghi, J., 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. Soc. Sci. Med. (1982), 285, 114277. https://doi.org/10.1016/j.socscimed.2021.114277.

Checkland, P., 1989. Soft systems methodology*. Hum. Syst. Manag. 8 (4), 273–289. https://doi.org/10.3233/HSM-1989-8405.

Checkland, P., Poulter, J., 2020. Soft systems methodology (Retired). In: Reynolds, En M., Holwell, S. (Eds.), Systems Approaches to Making Change: A Practical Guide. Springer, pp. 201–253. https://doi.org/10.1007/978-1-4471-7472-1_5.

Churchman, C.W., Ackoff, R.L., Arnoff, E.L., 1957. Introduction to Operations Research.

Crielaard, L., Uleman, J.F., Châtel, B.D.L., Epskamp, S., Sloot, P.M.A., Quax, R., 2022. Refining the causal loop diagram: A tutorial for maximizing the contribution of domain expertise in computational system dynamics modeling. Psychol. Methods. https://doi.org/10.1037/met0000484.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., Pappas, G., 2008. Comparison of PubMed, scopus, web of science, and Google scholar: strengths and weaknesses. FASEB J. 22 (2), 338–342. https://doi.org/10.1096/fj.07-9492LSF.

Folke, C., Polasky, S., Rockström, J., Galaz, V., Westley, F., Lamont, M., Scheffer, M., Österblom, H., Carpenter, S.R., Chapin, F.S., Seto, K.C., Weber, E.U., Crona, B.I., Daily, G.C., Dasgupta, P., Gaffney, O., Gordon, L.J., Hoff, H., Levin, S.A., Walker, B. H., 2021. Our future in the Anthropocene biosphere. Ambio 50 (4), 834–869. https://doi.org/10.1007/s13280-021-01544-8.

Forrester, J.W., 1961. Industrial Dynamics. MIT Press, Cambridge, MA.

Forrester, J.W., 1994. System dynamics, systems thinking, and soft OR. Syst. Dyn. Rev. 10 (2–3), 245–256. https://doi.org/10.1002/sdr.4260100211.

Forrester, J.W., Senge, P.M., 1980. Tests for Building Confidence in System Dynamics Models. System dynamics (MIT Press).

Funtowicz, Ravetz, J., 1997. Environmental problems, post-normal science, and extended peer communities. In: Études et Recherches Sur Les Systèmes Agraires et Le Développement, 169.

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. Ecol. Model. 198 (1–2), 115–126. https://doi.org/10.1016/j.ecolmodel.2006.04.023.

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. J. Artif. Soc. Soc. Simul. 23 (2), 7. https://doi.org/10.18564/jasss.4259.

Harzing, A.-W., Alakangas, S., 2016. Google scholar, scopus and the web of science: A longitudinal and cross-disciplinary comparison. Scientometrics 106 (2), 787–804. https://doi.org/10.1007/s11192-015-1798-9.

Jebb, A.T., Ng, V., Tay, L., 2021. A review of key likert scale development advances: 1995–2019. Front. Psychol. 12 https://www.frontiersin.org/articles/10.3389/fpsyg.2021.637547.

Kapsalis, T.Ap., 2022. A three-leg cultural relay race to sustainability of rural communities. Curr. Res. Environ. Sustain. 4, 100136. https://doi.org/10.1016/j.crsust.2022.100136.

Kinzig, A.P., Ehrlich, P.R., Alston, L.J., Arrow, K., Barrett, S., Buchman, T.G., Daily, G.C., Levin, B., Levin, S., Oppenheimer, M., Ostrom, E., Saari, D., 2013. Social norms and global environmental challenges: the complex interaction of behaviors, values, and policy. BioScience 63 (3), 164–175. https://doi.org/10.1525/bio.2013.63.3.5.

Lane, D.C., 1995. The Folding Star: A Comparative Reframing and Extension of Validity Concepts in System Dynamics.

Lane, D.C., Smart, C., 1994. Mad, bad and dangerous to know? The evolution and limitations of generic structure in system dynamics. In: Proceedings of the System Dynamics Society International Conference, pp. 67–77.

Levins, R., 1974. Discussion paper: the qualitative analysis of partially specified systems. Ann. N. Y. Acad. Sci. 231 (1), 123–138. https://doi.org/10.1111/j.1749-6632.1974.tb20562.x.

Liu, J., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Pell, A.N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C.L., Schneider, S.H., Taylor, W.W., 2007. Complexity of coupled human and natural systems. Science 317 (5844), 1513–1516. https://doi.org/10.1126/science.1144004.

Meadows, D. (1980). The unavoidable a priori. En J. Randers (Ed.), Elements of the System Dynamics Method. Pegasus Comunication.

Mousavi, S.H., Kavianpour, M.R., Alcaraz, J.L.G., 2023. The impacts of dumping sites on the marine environment: A system dynamics approach. Appl Water Sci 13 (5), 109. https://doi.org/10.1007/s13201-023-01910-9.

Nash, K.I., Cvitanovic, C., Ea, F., Bs, H., Ej, M.-G., Ra, W., Jl, B., 2017. Planetary boundaries for a blue planet. Nat. Ecol. Evol. 1 (11). https://doi.org/10.1038/s41559-017-0319-z.

Oliveira, B.M., Boumans, R., Fath, B.D., Harari, J., 2022a. Coastal ecosystem services and climate change: case study for integrated modeling and valuation. Global Ecol. Conserv. 38, e02240. https://doi.org/10.1016/j.gecco.2022.e02240.

Oliveira, B.M., Boumans, R., Fath, B.D., Harari, J., 2022b. Socio-ecological systems modelling of coastal urban area under a changing climate – case study for Ubatuba, Brazil. Ecol. Model. 468, 109953. https://doi.org/10.1016/j.ecolmodel.2022.109953.

Oliveira, B.M., Boumans, R., Fath, B.D., Harari, J., 2024. How social-ecological systems resilience unfolds from distinct worldviews. Front. Sustain. Res. Manag. 3. https://doi.org/10.3389/fsrma.2024.1352707.

Oliver, T.H., Benini, L., Borja, A., Dupont, C., Doherty, B., Grodzińska-Jurczak, M., Iglesias, A., Jordan, A., Kass, G., Lung, T., Maguire, C., McGonigle, D., Mickwitz, P., Spangenberg, J.H., Tarrason, L., 2021. Knowledge architecture for the wise governance of sustainability transitions. Environ. Sci. Pol. 126, 152–163. https://doi.org/10.1016/j.envsci.2021.09.025.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263 (5147), 641–646. https://doi.org/10.1126/science.263.5147.641.

Ostrom, E., 2000. Collective action and the evolution of social norms. J. Econ. Perspect. 14 (3), 137–158.

Ostrom, E., 2009. A general framework for analyzing sustainability of social-ecological systems. Science 325 (5939), 419–422. https://doi.org/10.1126/science.1172133.

Overmars, K.P., de Groot, W.T., Huigen, M.G., 2007. Comparing inductive and deductive modeling of land use decisions: principles, a model and an illustration from the Philippines. Hum. Ecol. 35, 439–452.

Parretti, P., Canning-Clode, J., Mendes, A.B., Costa, A.C., 2020. Who, why and how: stakeholder attitudes toward marine non-indigenous species management in Portuguese Atlantic Islands. Ocean Coast. Manag. 188, 105069. https://doi.org/10.1016/j.ocecoaman.2019.105069.

Payo, A., Hall, J.W., French, J., Sutherland, J., van Maanen, B., Nicholls, R.J., Reeve, D. E., 2016. Causal loop analysis of coastal geomorphological systems. Geomorphology 256, 36–48. https://doi.org/10.1016/j.geomorph.2015.07.048.

Pinnegar, J.K., Hamon, K.G., Kreiss, C.M., Tabeau, A., Rybicki, S., Papathanasopoulou, E., Engelhard, G.H., Eddy, T.D., Peck, M.A., 2021. Future socio-political scenarios for aquatic resources in Europe: A common framework based on shared-socioeconomic-pathways (SSPs). Front. Mar. Sci. 7 https://www.frontiersin.org/articles/10.3389/fmars.2020.568219.

Predragovic, M., Assis, J., Sumaila, U.R., Gonçalves, J.M.S., Cvitanovic, C., Horta e Costa, B., 2024. Up to 80% of threatened and commercial species across European marine protected areas face novel climates under high emission scenario. Npj Ocean Sustain. 3 (1), 1–9. https://doi.org/10.1038/s44183-024-00068-4.

Rahmandad, H., Lim, T.Y., Sterman, J., 2021. Behavioral dynamics of COVID-19: estimating underreporting, multiple waves, and adherence fatigue across 92 nations. Syst. Dyn. Rev. 37 (1), 5–31. https://doi.org/10.1002/sdr.1673.

Ravetz, J.R., 1993. The sin of science: ignorance of ignorance. Knowledge 15 (2), 157–165. https://doi.org/10.1177/107554709301500203.

Reed, M.S., Evely, A.C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., Stringer, L.C., 2010. What is social learning? Ecol. Soc. 15 (4). https://www.jstor.org/stable/26268235.

Reisman, A., Oral, M., 2005. Soft systems methodology: A context within a 50-year retrospective of OR/MS. Interfaces 35 (2), 164–178.

Renzi, M., Perra, G., Lobianco, A., Mari, E., Guerranti, C., Specchiulli, A., Pepi, M., Focardi, S., 2010. Environmental quality assessment of the marine reserves of the Tuscan archipelago, Central Tyrrhenian Sea (Italy). Chem. Ecol. 26 (sup1), 299–317. https://doi.org/10.1080/02757541003627647.

Richardson, G.P., 1986. Problems with causal-loop diagrams. Syst. Dyn. Rev. 2 (2), 158–170. https://doi.org/10.1002/sdr.4260020207.

Richardson, G.P., 1997. Problems in causal loop diagrams revisited. Syst. Dyn. Rev. 13 (3), 247–252. https://doi.org/10.1002/(SICI)1099-1727(199723)13:3<247::AID-SDR128>3.0.CO;2-9.

Richmond, B., 1994. Systems thinking/system dynamics: Let's just get on with it. Syst. Dyn. Rev. 10 (2–3), 135–157. https://onlinelibrary.wiley.com/doi/abs/10.1002/sdr.4260100204.

Rigét, F., Mosbech, A., Boertmann, D., Wegeberg, S., Merkel, F., Aastrup, P., Christensen, T., Ugarte, F., Hedeholm, R., Fritt-Rasmussen, J., 2019. The seas around Greenland: an environmental status and future perspective. World Seas: Environ. Evaluat. 45–68.

Ripple, W.J., Wolf, C., Newsome, T.M., Barnard, P., Moomaw, W.R., Grandcolas, P., 2019. World scientists' warning of a climate emergency. BioScience 70 (1), 8–12.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Foley, J.A., 2009. A safe operating space for humanity. Nature 461 (7263). https://doi.org/10.1038/461472a. Article 7263.

Sambolino, A., Alves, F., Fernandez, M., Krakauer, A.B., Ferreira, R., Dinis, A., 2022. Spatial and temporal characterization of the exposure of island-associated cetacean populations to whale-watching in Madeira Island (NE Atlantic). Reg. Stud. Mar. Sci. 49, 102084. https://doi.org/10.1016/j.rsma.2021.102084.

Sanò, M., Richards, R., Medina, R., 2014. A participatory approach for system conceptualization and analysis applied to coastal management in Egypt. Environ. Model Softw. 54, 142–152. https://doi.org/10.1016/j.envsoft.2013.12.009.

Schadeberg, A., van Leeuwen, J., Groeneveld, R.A., Kraan, M., 2024. Science is not enough: the role of legitimacy in the governance of marine activities. Mar. Policy 169, 106337. https://doi.org/10.1016/j.marpol.2024.106337.

Schwaninger, M., Groesser, S., 2020. System dynamics modeling: Validation for quality assurance. In: Dangerfield, En B. (Ed.), System Dynamics: Theory and Applications. Springer US, pp. 119–138. https://doi.org/10.1007/978-1-4939-8790-0_540.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., Ludwig, C., 2015. The trajectory of the Anthropocene: the great acceleration. Anthropocene Rev. 2 (1), 81–98.

Sterman, 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World.

Sterman, J.D., 2002. All models are wrong: reflections on becoming a systems scientist. Syst. Dyn. Rev. 18 (4), 501–531. https://doi.org/10.1002/sdr.261.

Van den Belt, M., 2004. Mediated Modeling: A System Dynamics Approach to Environmental Consensus Building. Island press.

Walker, B., Crépin, A.S., Nyström, M., Anderies, J.M., Andersson, E., Elmqvist, T., Queiroz, C., Barrett, S., Bennett, E., Cardenas, J.C., Carpenter, S.R., Chapin, F.S., de Zeeuw, A., Fischer, J., Folke, C., Levin, S., Nyborg, K., Polasky, S., Segerson, K., Vincent, J.R., 2023. Response diversity as a sustainability strategy. Nat. Sustain. https://doi.org/10.1038/s41893-022-01048-7.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A.G., De Souza Dias, B.F., Ezeh, A., Frumkin, H., Gong, P., Head, P., Horton, R., Mace, G.M., Marten, R., Myers, S.S., Nishtar, S., Osofsky, S.A., Pattanayak, S.K., Pongsiri, M.J., Romanelli, C., Yach, D., 2015. Safeguarding human health in the Anthropocene epoch: Report of The Rockefeller Foundation–Lancet Commission on planetary health. Lancet 386 (10007), 1973–2028. https://doi.org/10.1016/S0140-6736(15)60901-1.