MARINE SABRES

Funded by the European Union

Marine SABRES Deliverable 4.1

Baseline assessment of Social-Ecological System models

DOCUMENT INFORMATION

Version	Date	Description			
		Responsible	Authors	Reviewed by	Approved by
1	28/08/2023	WP4	Angel Borja Bruno Meirelles de Oliveira		
2	20/10/2023	WP4	Angel Borja Bruno Meirelles de Oliveira		
3	29/02/2024	WP4	Those listed below	Herman Hummel	
4	5/03/2024	WP4	Those listed below	Arturas Razinkovas, Herman Hummel Zacharoula Kyriazi	Emma Verling
5	7/06/2024	WP4	Those listed below		Emma Verling
6	4/07/2024	WP4	Those listed below		Emma Verling

Authors	
Name	**Organisation**
Angel Borja	AZTI
Bruno Meirelles de Oliveira	AZTI
Sarai Pouso	AZTI
Ana Cristina Matos Ricardo Costa	University of Azores
Andrea Zita Costa Botelho	University of Azores
Gustavo Oliveira de Meneses Martins	University of Azores
Manuela I. Parente	University of Azores
Anunciação Ventura	University of Azores
Nuno Álvaro	University of Azores
Ana Diniz	ARDITI
João Canning-Clode	ARDITI
Paola Parretti	ARDITI
Furqan Asif	Aalborg University
Josefin Ekstedt	Aalborg University
Anna Heiða Ólafsdóttir	Marine and Freshwater Research Institute
Bjarki Þ. Elvarsson	Marine and Freshwater Research Institute
Sandra Rybicki	Marine and Freshwater Research Institute
Pamela J. Woods	Marine and Freshwater Research Institute
Mirjam Carlsdóttir Olsen	Blue Resource

Unn Laksá	Blue Resource
Talea Weissang	WWF Greenland
Caterina Mintrone	University of Pisa
Lisandro Benedetti-Cecchi	University of Pisa
Gemma Smith	IECS Ltd.
Michael Elliott	IECS Ltd.

Acknowledgements/contributions	
Name	Organisation
João Gama Monteiro	ARDITI
Patrício Ramalhosa	ARDITI
Rodrigo Silva	ARDITI
Silvia Almeida	ARDITI
Susanne Schäfer	ARDITI
Pedro Afonso	IMAR – Instituto do Mar
Sofia Garcia	DRAM - Direção Regional de Políticas Marítimas
Maria Dulce de Oliveira Resendes	Capitania do Porto de Vila do Porto
Internal reviewers	
Heman Hummel	HUFOSS
Michael Elliott	IECS Ltd.
Emma Verling	University College Cork
Zacharoula Kyriazi	University College Cork
Arturas Razinkovas	Klaipeda University

DISCLAIMER

How to cite this report: Borja, A., Oliveira, B., Pouso, S., A.C. Matos, A.C. Costa, A.Z. Botelho, J.M.N. Azevedo, G. M. Martins, M.A. Ventura, M.I. Parente, N. Alvaro, A. Diniz, J. Canning-Clode, P. Parretti, F. Asif, J. Ekstedt, A.H. Olafsdóttir, B. Elvarsson, S. Rybicki, P.J. Woods, M. C. Olsen, U. Laksá, C. Mintrone, L. Benedetti-Cecchi, G. Smith, M. Elliott, 2024. Baseline assessment of Social-Ecological System models. Marine SABRES Deliverable 4.1, 171 pages

This deliverable is pending official approval by the European Commission and therefore the contents are solely the opinion of the authors and not of the European Commission.

Table Of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Complete name
AoS	Appreciation of the Situation
BOT	Behaviour Over Time
CCM	Communicated Conceptual Model
CLD	Causal Loop Diagram
D	Deliverable
DA	Demonstration Area
EBM	Ecosystem-Based Management
EEA	European Economic Area
EEZ	Exclusive Economic Zone
EU	European Union
FAMRI	Faroe Marine Research Institute
FM	Formal Model
GDP	Gross Domestic Product
GVA	Gross Value Added
ICES	International Council for the Exploration of the Sea
ISA	Integrated System Approach
ISTAT	Italian National Institute of Statistics
ITQ	Individual Transferable Quotas
MFRI	The Marine and Freshwater Institute of Iceland
MPA	Marine Protected Area
NEAFC	Northeast Atlantic Fisheries Commission
NGO	Non-Governmental Organizations
PIMS	Process and Information Management System
PIoR	Policy Insights or Recommendations
SES	Social-Ecological Systems
SFPA	Sustainable Fisheries Partnership Agreement
sSES	Simple Social-Ecological Systems
SME	Small and Medium Enterprises
SSB	Spawning Stock Biomass
SSM	Soft Systems Methodologies
SST	Sea Surface Temperature
T	Task
TAC	Total Allowable Catch
TANP	Tuscan Archipelago National Park
WP	Work Package

Executive Summary

This baseline assessment is the departure point of the application of the simple socio-ecological systems (SES) approach to each Demonstration Areas (DAs). This has been done by testing the method for studying and analysing marine SES, which was developed in WP3 and is called the Simple SES (sSES) (Sections 5-9). For enhancing the robustness and usability of the outcomes of the first application of the sSES, the baseline assessment process (which is part of Task 4.1 as described in the project's DoA) was enriched and extended to include the development and application of an additional modelling approach (Section 10). This was suggested by the WP4 lead and agreed upon by the project consortium, after the first application. The three DAs (Tuscany, Macaronesia and Arctic) were involved in the whole testing process (including both methods) and are responsible for the content in each resulting model of their respective SES, which represent their knowledge of the issues of each DA (i.e., tourism and seagrass meadows, tourism and the ecological corridor, and pelagic fisheries, respectively).

This deliverable consists of Sections 1 and 2, which include the description of the project and a statement of the purpose of the baseline assessments of the DAs SES, in accordance with the suggested extended process as described above. Sections 3-8 describe the first iteration of modelling, including the description of the DAs and how the sSES works, followed by the results of its application, its validation, and a discussion of the process. Conclusions and recommendations, in Section 9, present the main issues with the tool, revealed by the previous sections, and provide recommendations for its refinement.

As the results of the first iteration were not considered sufficiently robust to represent the situation of the DAs and to proceed with the other steps of the project, Section 10 includes the description of the suggested additional modelling approach, which was applied in the DAs while facilitated by the WP4 lead. The outcomes, conclusions and recommendations of this new baseline assessment are provided, connecting the results of this iteration with the other WPs in Marine SABRES.

As shown by the baseline assessment, each DA faces unique challenges influenced by local socio-economic contexts and environmental pressures, emphasizing the need for tailored management strategies.

In the Tuscany DA, tourism was identified as a crucial driver of economic prosperity, but also for environmental stress on marine ecosystems. The study of this DA's SES aimed to understand tourism interactions with other elements of the SES, especially its interactions with the seagrass meadows, and ways to promote marine conservation and restoration.

In the Macaronesia DA, the tourism interaction with marine protected areas was explored and the possibility of a marine corridor, connecting the three archipelagos was discussed. The analysis revealed the tourism dual role as the cause of a pressure on ecosystems and a provider of socioeconomic benefits.

In the Arctic DA, the focus was on the impact of commercial pelagic fisheries in the Northeast Atlantic, in the Faroe Islands, Iceland, and Eastern Greenland in an integrated model, that revealed common issues and challenges for this region.

For both the sSES and the new iteration approach, a testing and validation process were then undertaken based on a guidance protocol developed as part of Task 4.1 for conducting a presumed utility of causal loop analysis, that includes "running" 26 tests that are categorized into four groups: broad guidelines and processes, specific model tests, policy insights, and administrative and overview reviews. The outcome of this process is a mosaic of positive, neutral, "not applied" and negative evaluations for each group, and iteration. For the first application of the sSES, some of the most positive aspects, such as the previous stakeholder interaction with researchers, the documentation of the modelling process, its purpose, and the meaning of the whole process reinforce the potential relevance of the sSES as a tool for the integrated analysis of coupled human-nature systems in coastal areas. Nonetheless, the negative evaluations and comments provided the substance to conclude the necessity to develop and apply the additional modelling approach and to indicate where refinement of the sSES is required, to reach the final goals of the modelling process. These comments also showed the importance, as described in the WP3 guidance document, of one or more subsequent iterations of the sSES. **The conclusions drawn from the first application of the sSES** highlight several important insights (not exhaustive):

- The need for better knowledge and/or explanation of **systems science**, regarding the assumptions and limitations of the field and particularly of the sSES, exploiting broad concepts, but also on relevant differences from traditional modelling techniques. In addition, how to integrate and deal with plural topics of the SES such as incompatible datasets, different governance regimes, multi-scale governance and others.

- **Data Scarcity and Quality**: All three DAs encountered challenges related to data scarcity and data quality as well as guidance in dealing with data-poor relevant topics. The lack of long-term, high-resolution datasets tailored to specific regions posed significant obstacles. The exclusion of relevant indicators due to the lack of data and issues with variable links were common.

- **The complexity of modelling using the sSES**: The manual process of updating the models with new elements or connections was identified as time-consuming and inefficient. While expert judgment might play a crucial role in reconciling discrepancies between quantitative data trends and loops, establishing connections based on Behaviour over Time (BOT) was problematical as it conflicted with expert knowledge.

- **Stakeholder use and broader communication**: Clear guidelines for model structure, data integration, and interpretation are essential for accurate representation and policy relevance. Improved communication strategies are needed to translate insights into actionable policies and gain community acceptance.

- **Challenges in Loop Analysis**: The Loop analysis revealed challenges including difficulties in establishing causal relationships, interpreting loop polarity, and ensuring model reproducibility. Hence, there is a need for clearer guidelines and formal analytical procedures to enhance the applicability and effectiveness of the sSES.

- **Validation and Future Development**: Stakeholder validation should be considered in future iterations of sSES modelling to ensure that major components of system function are captured, trust in the results is promoted, and robust decisions are supported.

In conclusion, the application of the sSES in the Tuscany Archipelago, Macaronesia, and Arctic DAs, has provided valuable insights into the complexities of social-ecological systems and the challenges and opportunities associated with modelling and decision-making in these contexts, but yet not robust enough to sustain the next steps of the project, what justifies the creation of the new iteration process.

The results produced in the new iteration are considered robust, meaningful, and integral to the Marine SABRES goals, which is novel and timely. To each DA, four categories of results were presented, namely: a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol. **These four types of results complement and integrate each other and are integral parts of the baseline assessment of each DA's SES as required by the DoA of Marine SABRES to be present in D4.1. WP4 recommends the consortia to use data from the Section 10 to proceed with the project**

1 Project Overview

The main goal of Marine SABRES is to contribute to European-wide efforts for biodiversity conservation by integrating sustainable marine ecosystems and a resilient blue economy. Marine SABRES is a four-year Horizon Europe research project aimed at enabling marine and coastal managers to make sustainable decisions, empowering citizens to engage with marine biodiversity conservation, promoting sustainable development in coastal and marine sectors, and setting European marine management on a course to reverse biodiversity decline.

In pursuit of this aim, Marine SABRES brings a diverse group of stakeholders from government, policy, and business and coastal management together with marine scientists to co-design a simple Social-Ecological System (sSES) approach. This approach will accelerate the uptake of Ecosystem-Based Management (EBM) to strengthen interventions and measures for the protection and conservation of coastal and marine areas and their biodiversity and thus to safeguard the undisturbed provision of Ecosystem Services and Societal Goods and Benefits.

Marine SABRES brings together an interdisciplinary team of leading international experts in social science, marine governance, stakeholder engagement, marine ecology and ecosystem-based management, environmental and ecological economics, and science communication. Our consortium is made of 21 partner organisations from eleven European countries, and is balanced with representatives from academia, government agencies, SMEs, and NGOs. Marine SABRES is coordinated by University College Cork, Ireland at MaREI, the SFI Centre for Energy Climate and Marine.

2 Purpose of the Baseline Assessment

The main objective of WP4 is to co-develop, test, and demonstrate the sSES approach, i.e. the marine social-ecological system analysis framework created by WP3 (Gregory et al., 2023). This main objective is separated into four operational objectives:

(i) Testing and demonstrating the sSES approach. This objective should be completed in Task T4.1, starting in December 2022, and ending in March 2024, with Deliverable D4.1 (Baseline assessment of SES), i.e. the present document; As an additional modelling approach was required, the deliverable was delayed to July 2024.

(ii) Developing and implementing bespoke conservation and restoration interventions;

(iii) Developing pathways and options for interventions. These two objectives should be completed in T4.2, which will extend from April 2024 to February 2025, with D4.2 (Options and pathways report); and

(iv) Testing and applying governance solutions. This objective will be completed in T4.3, starting in February 2025 and ending in January 2026, with D4.3 (Option appraisals).

The above aspects will be undertaken in a cyclical three-step process, which will be conducted in each of the three Demonstration Areas (DAs; Tuscany Archipelago, Macaronesia and Arctic) to enable the implementation of biodiversity conservation measures and by co-developing and testing the sSES its guidance and its supporting tools. Stakeholders are involved in all steps of the process:

(i) To analyse baseline system function using the sSES and supporting tools, identify and trial measures. The outcome of the sSES application was complemented with a new iteration to proceed with steps ii and iii

(ii) To identify pathways for transformation and solutions and further measures for management.

(iii) To set goals and objectives (based on stakeholder and policy objectives).

Within this framework, Task 4.1 was about describing the baseline system function in each of the DAs. Where necessary, location-specific parameters, indicators and proxies would be selected for each model subsystem, the links between them would be characterised and quantified.

The baseline assessments of each DA's SES that was attempted by the application of the sSES are presented in Section 7. The complementary baseline assessments as resulted by the new iteration are presented in Section 10.

As the results of the later are considered more robust and refer to each DA's SESs as a whole (and not to the separate countries comprising them as it was the case with the sSES application), they can be used already for an initial appraisal of the situation in the DAs. Some exploratory navigation of the systemic properties of each variable might be profitable for management, considering the initial maturity stage of the model. The results of this baseline assessment, especially the loops and the variables of interest (described in Section

10), can deepen the systems analysis using stakeholder inputs as complementary knowledge. The fact that distinct groups often see the systems in different ways and frame them in different model structures is trivial knowledge in systems analysis, it is emphasized there is little sense in comparing the structure of the CLDs produced with a possible structure made by stakeholders (in T2.2). Therefore, it is suggested that the CLDs produced by applying the new iteration can be used to communicate the expert views of the system structure and then to be complemented punctually inside a small set of loop analyses, by specific knowledge from the stakeholders groups. This approach can complement the results presented here, ensuring they capture the major components of the examined SES function and are sufficiently robust to support decisions (Task 2.3). Initial concrete management measures are identified in the DA descriptions, these are the starting point for further concrete measures development and trialling in each DA.

3 Overview of the Demonstration Areas

This chapter presents a description of each of the DAs, for a better understanding of the three areas in which the sSES is going to be tested. This helps to better appreciate the context regarding the development of the Marine SABRES approach.

3.1 Tuscany

Geographical description

The Tuscan Archipelago National Park (TANP) is one of the largest marine parks in Europe and it has been declared a Biosphere Reserve by UNESCO in recognition of its unique environments and the large diversity of marine and terrestrial life that it supports. The TANP was established in 1996 and spans 614.7 km^2 of sea between the Ligurian and the Tyrrhenian Seas; it currently represents 20% of the marine protected areas in Italy (Figure 1). The TANP includes seven islands that are managed differently concerning human activities, with restrictions ranging from fully protected islands, such as Montecristo and Pianosa, with limited human access and no extractive activities, to islands that alternate marine protected areas (such as Capraia island) with open areas where human access and recreational activities are allowed, to fully open islands (Elba).

The area includes 10 Municipalities and 2 Provinces. Of the seven main islands, Elba, Giglio and Capraia have a resident population throughout the year, which approximates to a total of 33,500 inhabitants during winter, whereas the summer population increases to 200,000 people occupying the islands, an increase largely due to tourism and, to a lesser extent, to residents that spend winters on the mainland and move to the islands in summer.

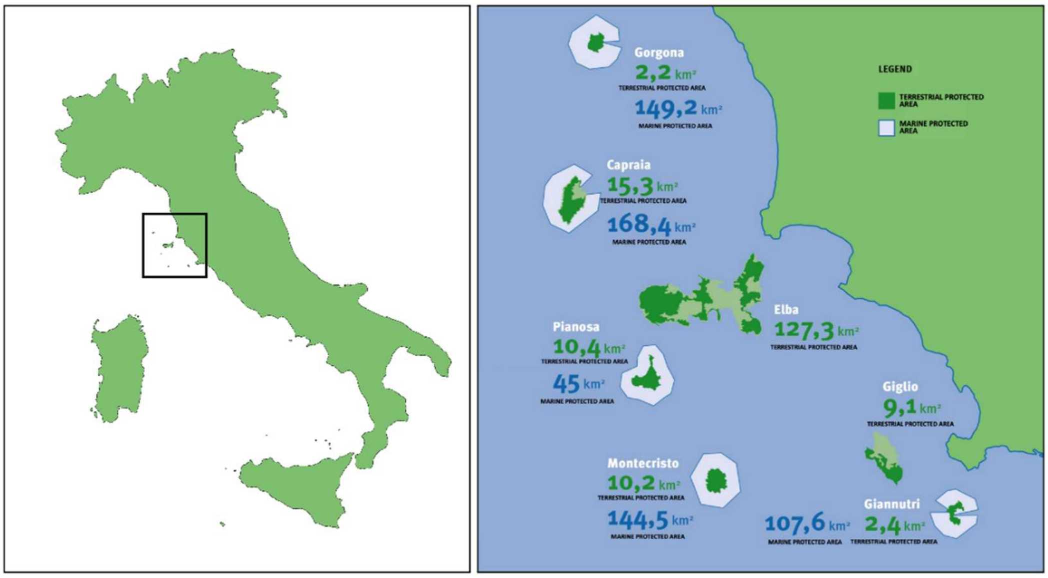

Figure 1. Map of Italy showing the location of the Tuscan Archipelago. On the right a detailed map shows the extension of Tuscan Archipelago National Park.

Biological description

Posidonia oceanica (L.) Delile meadows are one of the most productive and widespread marine coastal ecosystems in TANP. Seagrass meadows provide essential ecosystem services (Duffy et al., 2019), such as the provision of nursery areas for commercially important species (Unsworth et al., 2019) and regulating services such as nutrient cycling (McGlathery et al., 2007), sediment stabilization (Barcelona et al., 2021; Granata et al., 2001) and significant carbon sequestration (Duarte et al., 2005). Seagrasses are particularly vulnerable to degraded environmental conditions and can collapse into an alternative state of algal turfs (low-lying, small-form algae), with cascading impacts on productivity, de-oxygenation, and ecosystem services (Montefalcone et al., 2015; Telesca et al., 2015; Waycott et al., 2009).

Economy and Environmental Pressures

The economy of the TANP is primarily driven by tourism. In 2021, 460,000 tourists visited the area for a total of 2,764,000 presences (number of nights spent at a given accommodation facility). Both arrivals and presences have increased in the last decade. However, tourists tend to spend progressively fewer nights on the islands, which reduces the economic impacts and benefits on local commercial activities. Tourism peaks between June and September (a period in which 85% of tourists are recorded). Spring (April-May) and autumn (October) represent the low season, with the winter months recording almost no tourists. The islands are attractive locations for sailing boats and yachts (Figure 2). Recreational diving activities are promoted through diving centres and boat rentals, both on the islands and the mainland.

Figure 2. Aerial view of the marina and village of Capraia.

Other economic activities include agriculture with high-quality agricultural production, consisting of grain, spirits, and wine products, artisanal fishing, and aquaculture. Capraia Island, for example, hosts an innovative aquaculture farm that has expanded over the last decade to include 8 circular cages each with a diameter of 25 m. This plant supplies 250 t of sea bream and sea bass per year, which are widely distributed across several markets on the mainland.

With more than two million visits each year, tourism provides the most important economic asset for the TANP, but it also causes measurable ecological impacts (Figure 3).

Figure 3. High density of boats anchored at Giannutri island.

Tourism drives important environmental changes, exerting pressure on vulnerable marine ecosystems, in particular, *Posidonia oceanica* meadows, through elevated nutrient discharge, disturbance (e.g., anchoring), and pollution (Figure 4). Tourism pressure is likely to increase in the coming years owing to global warming, with the holiday season lasting longer (Perry, 2000; Rutty and Scott, 2010).

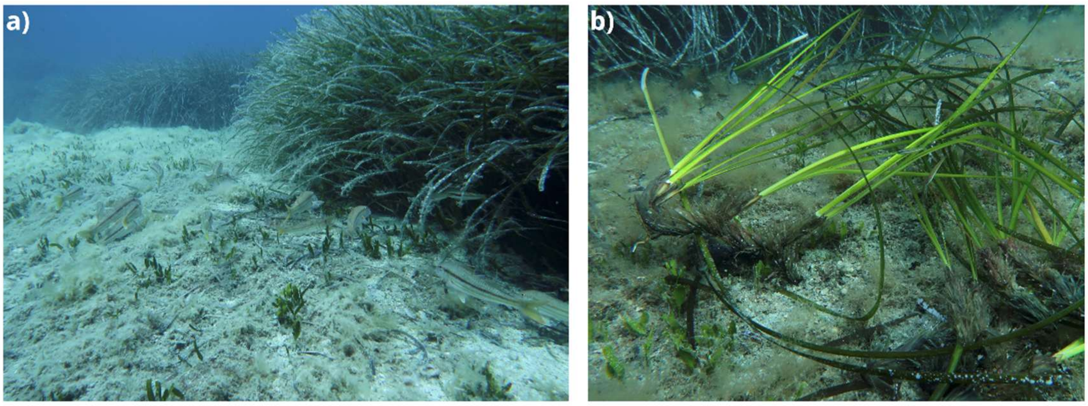

Figure 4. Posidonia oceanica meadow impacted by anchorages at Giannutri island. Panel a) shows the contrast between the meadow and adjacent turf-dominated area, panel b) shows rhizomes detached by an anchor.

Focus of this study

The broad vision of this DA is to balance tourism and the conservation of seagrass meadows in the TANP. By using the sSES approach, the DA participants modelled the main connections between tourism, society, and coastal marine ecosystems, with the ultimate aim of identifying effective management options for the conservation and protection of seagrasses and the ecosystem services they provide, as well as considering the requirements to adapt to climate change and the challenges of implementing Marine Protected Areas (MPAs) in a complex governance landscape. To advance the understanding of the current state and temporal dynamics of the system, we gathered and integrated many sources of information. First, we listed the major environmental laws and rules that are implemented in the TANP, together with the players in local and national governance. Socio-economic data for the TANP have been mainly derived from the Italian National Institute of Statistics (ISTAT) and Park Authority periodical reports. Environmental data have been mainly downloaded from the NOAA/OAR/ESRL PSD website, Boulder, Colorado, USA. Ecological data have been collected from multiple sources such as official reports of national monitoring programmes conducted by regional or national environmental agencies or observational data collected from the research group of the University of Pisa.

3.2 Macaronesia

Geographical description

Macaronesia is a biogeographic area in the NE Atlantic that includes the **Azores, Madeira**, and **Canary Islands** archipelagos, as well as, in a broader context, the **Cabo Verde** islands (Figure 5).

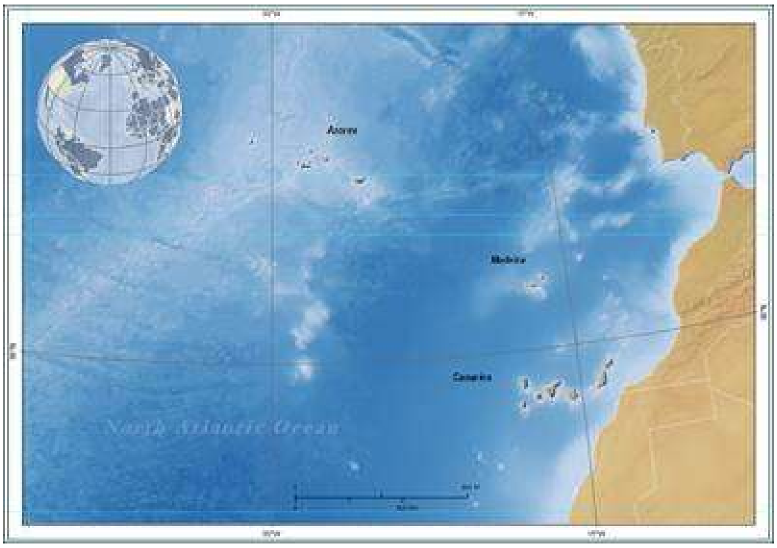

Figure 5. Macaronesia's archipelagos. FCT/UAc, 2024

In the proposal, Canary Islands were included to test the sSES together with Azores and Madeira. As for unexpected reasons, the partner of Canary was not able to apply the model in this archipelago, the activities herein reported were conducted only in the Portuguese Macaronesia biogeographic region: Azores and Madeira archipelagos. Both archipelagos are composed of volcanic islands (Azores 9; Madeira 2 main islands and two sub-archipelagos) and several islets. These islands cover a latitudinal range from 30°N to 39°N and a level of isolation (from the nearest continental point) that ranges between 520 km (Madeira Island) to +1,500 km (Flores, Azores). Madeira Island has a population density of almost 300 people km^{-2}, standing among the highest densities in the European Union. The Azores have the lowest population density in Macaronesia, with less than 60 people km^{-2} (Figure 6).

Figure 6. (Above) Funchal, the capital of Madeira Island, © Bernhard1960; (Bellow) Santa Cruz, the main village in Flores island, Azores archipelago © CMSCF,

Biological description

These islands support a diverse fauna and flora with high levels of taxonomic endemism in some groups, particularly from the terrestrial realm such as plants, snails and arthropods (Florencio et al., 2021). The Macaronesian archipelagos also encompass a wide range of marine coastal habitats, including seagrasses, wetlands and Vulnerable Marine Ecosystems (VMEs), and support a rich and abundant marine biota (Figure 7), including megafauna, pelagic fish (e.g. tuna and swordfish) and seabirds

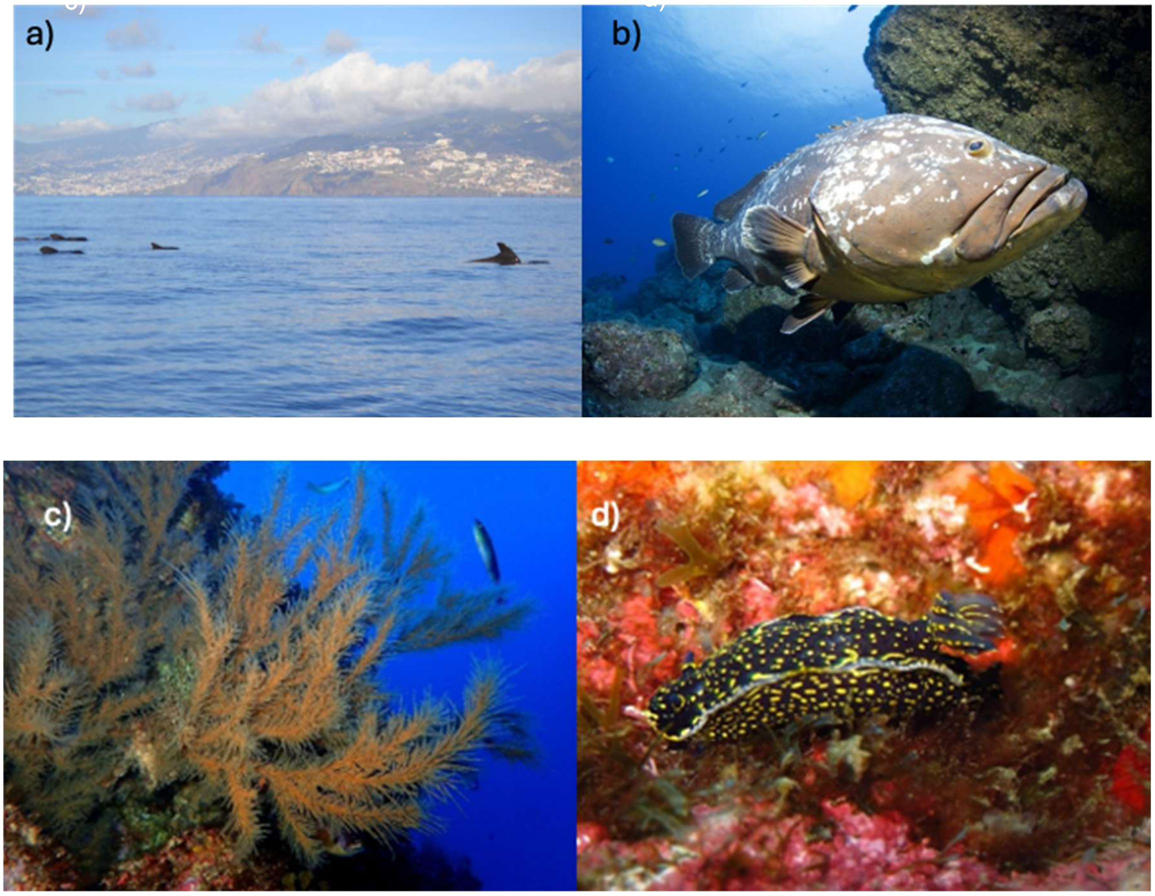

Figure 7. a) Short-finned pilot whales (Globicephala macrorhynchus), © F. Alves/MARE-Madeira; b) Dusky grouper (Epinephelus marginatus), © J. Monteiro/MARE-Madeira; c) Black coral (Antipathella wollastoni), ©Enric Ballesteros; d) Nudibranch (Felimare picta), ©CI

Economy and Environmental Pressures

The Macaronesia region's economy is strongly specialized in the services sector, where tourism has a prominent role, particularly in Madeira (two million visitors in 2023). Madeira archipelago has a long history of tourist activity, but it is still experiencing an increase, following the global trend of the last few years (Figure 8a). Fishing and agriculture also have an economic role (especially black scabbard fish fishery, and banana production) but are secondary compared to tourism. In the Azores, agriculture is economically important, with a significant predominance of livestock and dairy production, together with fisheries, although tourism has developed substantially in the last decade (Figure 8b). These growing activities exert pressure on the archipelagos and especially on the areas addressed in the present study.

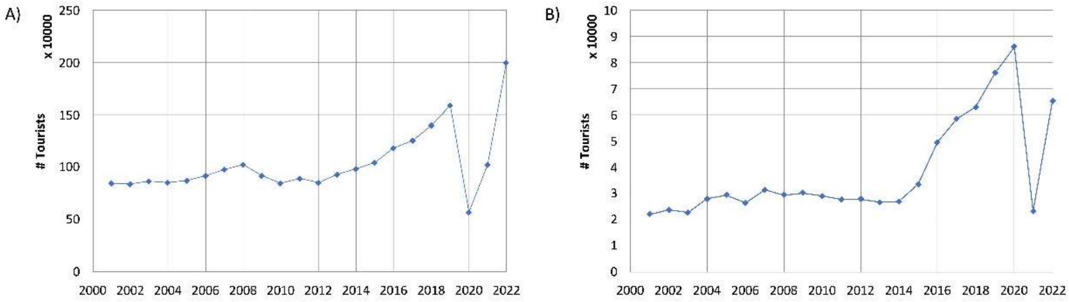

Figure 8. Number of tourists in the Portuguese islands of Macaronesia over the past 20 years. A) in the Madeira archipelago and B) in the archipelago of the Azores (data source: Direção Regional da Estatística da Madeira and Observatório de Turismo dos Açores).

Focus of this study

The balance between the economic and societal benefits of tourism and its ecological impact on Macaronesia coastal habitats (Figure 9) was identified as a major challenge in the region. We, therefore, modelled the primary connections between tourism, society and coastal marine ecosystems, with an emphasis on the conservation of MPAs. Data from different sources were gathered, including official national data (e.g., national and regional statistics entities), regional agencies (e.g., local governments) and data collected within the scope of marine research projects and monitoring programmes led by the University of Azores and MARE-ARDITI research centres. All data were analysed within the sSES approach, which aims to provide a holistic understanding of the system temporal dynamics, designed to identify connections and management options that would foster a sustainable balance between the services provided by marine ecosystems, their conservation, and the societal benefits of tourism.

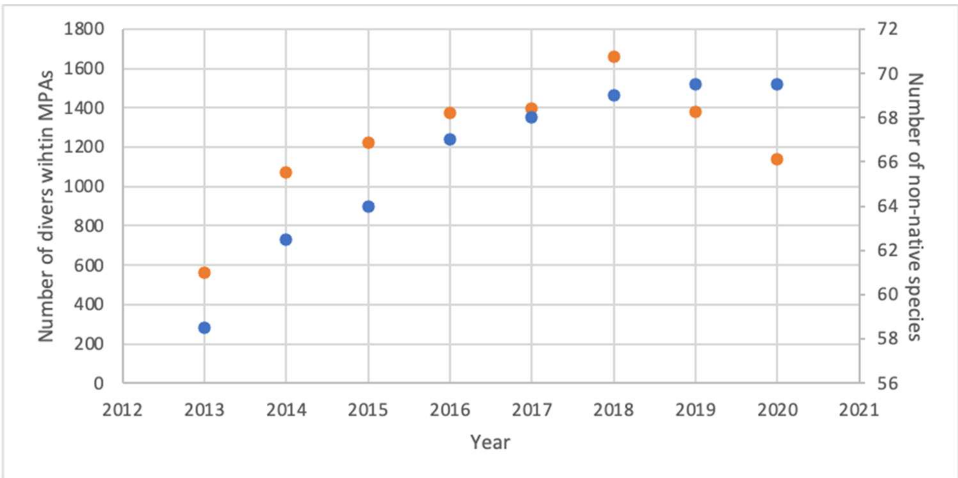

Figure 9. Temporal variation in the number of divers averaged across four MPAs (orange circles) and the accumulated number of non-native species (blue circles) (data source: Capitania do Porto de Vila do Porto & CIBIO-Açores).

3.3 Arctic

The Arctic DA consists of three countries: Greenland, Iceland, and the Faroe Islands (Figure 10). The continental shelf areas of these countries support diverse and productive ecosystems that play a central role in the national economies of the respective countries. They sit on a dynamic front between the cold Arctic Ocean current and the warmer Gulf current and incorporate the meltwater from Greenland glaciers. In recent years, fish species have undergone distribution shifts both within the region (e.g., haddock in Iceland) and to the region (e.g. mackerel), providing a unique set of management challenges related to both the social and environmental aspects of the system. The Faroe Islands, Iceland, and Greenland are closely interconnected through the economic structure of commercial fisheries, as they share ocean space, several commercially important fish stocks, heavy economic dependence on commercial fishing, and concerns about the impact of climate change on the ocean environment.

Figure 10. Arctic Demonstration Area (DA) encompassing Greenland, Iceland, and the Faroe Islands.

However, there are differences between the three nations in terms of how the fishing industry has developed (for example, the vertical structure of the companies), how fisheries are governed (licenses, quotas, bilateral agreements, etc.), and how fisheries are tied to cultural and economic aspects of their societies. In Marine SABRES, these three countries represent one international research area, but the countries have different environmental, economic, social, and governance settings. It is therefore important to consider the specific national contexts when working with the research sites. The sections below briefly describe the three countries.

Faroe Islands (description, economy and environmental pressures)

The Faroe Islands are an archipelago of 18 mountainous islands, with a population of 54,547 people (Statistics Faroe Islands, 2024), a total land area of 1,399 km^2 and 1,117 km of coastline (Government of the Faroe Islands, 2024). The EEZ sea area is 274,000 km^2 (Figure 11) (ICES, 2023c).

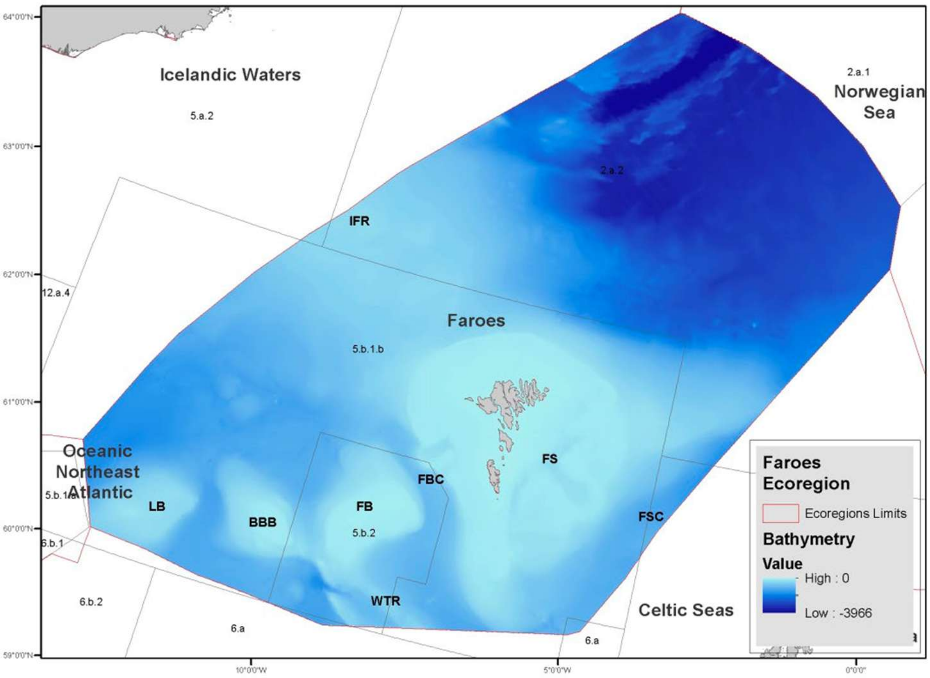

Figure 11. The Faroes ecoregion as defined by ICES. ICES areas are indicated by thin grey lines. Source: reproduced from ICES (2023c).

The Faroe Islands are a self-governing country associated with the Kingdom of Denmark. The country is a parliamentary democracy with a multi-party system, with the Prime Minister of the Faroe Islands being the head of government. The Faroe Islands have exclusive competence to legislate and govern independently in most areas. These include, for example, the conservation and management of living marine resources, protection of the environment, sub-surface resources, trade, taxation, industrial relations, energy, transport, communications, social security, culture, education, and research. It is of note that the country is not a member of the EU. Accordingly, the Faroe Islands negotiates its own trade and fisheries agreements with the EU and other countries and participates actively in a range of international fisheries management arrangements and organizations (Government of the Faroe Islands, 2024).

The Ministry of Fisheries is responsible for the management of fisheries and for the implementation of relevant legislation. The Ministry issues regulations for commercial fishing on an annual basis (every fishing year), allocating the number of days at sea and Total Allowable Catch (TAC) of stocks that are subject to limitations. The fisheries for some stocks are managed based on agreements between Northeast Atlantic Fisheries Commission (NEAFC) and coastal states and by bilateral agreements (ICES, 2023c). In particular, the pelagic fisheries for Norwegian spring-spawning herring, mackerel, and blue whiting are regulated by quotas according to coastal state agreements (ICES, 2023c). For demersal fisheries targeting cod, haddock, and saithe, an effort management system has been in place since 1996. The total effort was adjusted according to a newly introduced management plan in 2021, that considers the state of stocks. For both demersal and pelagic fisheries, fisheries advice is provided by the Faroe Marine Research Institute (FAMRI) and ICES.

In the Faroe Islands, employment in fisheries has gradually decreased despite the continuously high catches. This is both due to the modernisation of the fishing fleet, resulting in increased efficiency (i.e., need for less labour), as well as due to a decrease in demersal fisheries and fish processing plants. Nevertheless, employment in the fisheries and aquaculture sector still represents about 10% of the total workforce. Currently, 2,600 individuals are employed in the Faroese fishing industry which encompasses wild capture fisheries, fish processing, and a growing aquaculture sector (Figure 12). In particular, the sector is particularly important outside the capital where non-fishing-sector-related jobs are few (ICES, 2023c).

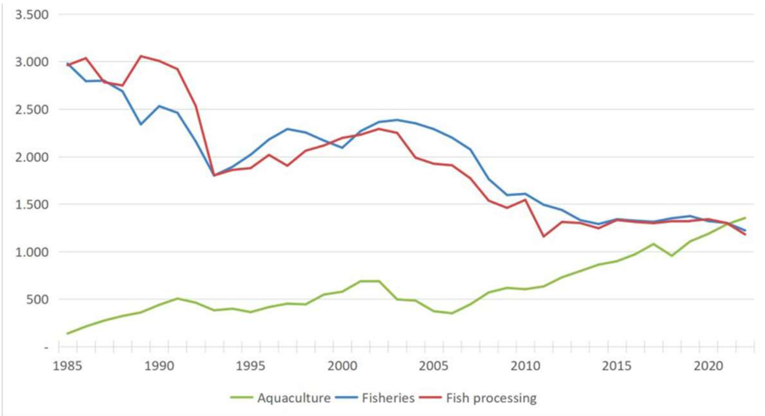

Figure 12. Number of employees in fisheries, aquaculture, and fish processing in Faroe Islands (Source: Statistics Faroe Islands, 2024)

An ecosystem overview for the Faroe Islands, recently published by ICES, includes an evaluation of activities and pressures relating to the area (Figure 13).

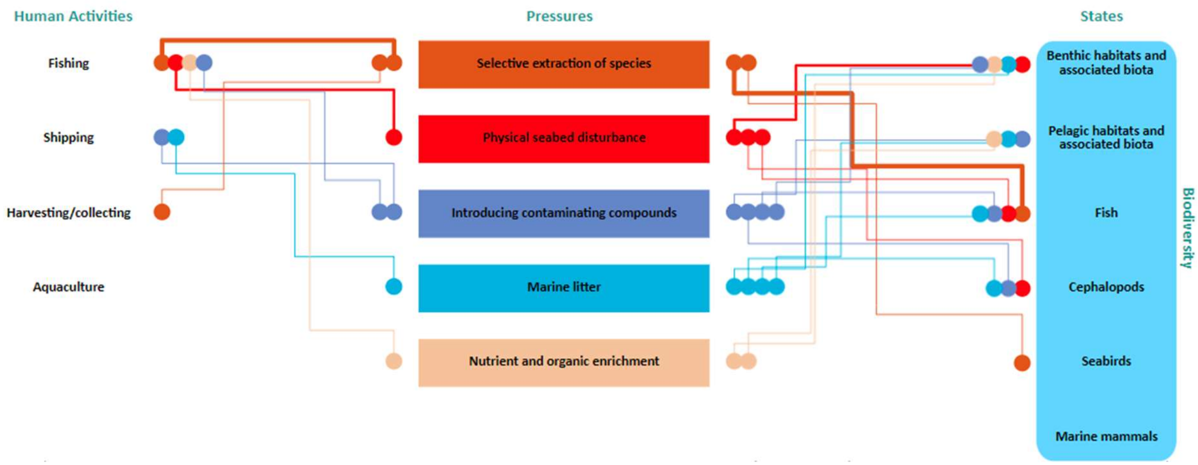

Figure 13. Major regional pressures, human activities, and ecosystem state components in Faroe Islands. The width of the lines indicates the relative importance of the main individual links. Each human activity and pressure is listed in decreasing order of its relative contribution to the total risk score. The absence of a line does not necessarily imply a total absence of any link; only the main links are shown. Climate change affects human activities, the intensity of the pressures, some aspects of the state, and the links between these Source: reproduced from ICES (2023c).

The overview shows that pressures caused by fisheries account for 76% of the impact of human activities on the ecosystem. Overall, selective extraction of species (43%), seabed disturbance (26%), and contaminants (11%) are the three highest pressures on the ecosystem, accounting for 80% of the total impact in the ecoregion (ICES, 2023c). In addition to fisheries, other dominating human activities are shipping, harvesting/collecting (e.g., marine mammal hunting and seaweed vegetation harvesting), and aquaculture. However, while fisheries occur almost everywhere in the ecoregion, shipping is mostly confined to fixed shipping routes. Similarly, harvesting/collecting and aquaculture are mainly limited to fjords and sounds in the islands.

Greenland (description, economy and environmental pressures)

Greenland is the world's largest island with a population of 50,160 people (Grønlands Statistik, 2023) and a total land area of 2,166,086 km^2, of which 80% is covered by inland ice and glaciers and a coastline covering 44,087 km (The Nordic Council, 2024). The total marine area is 2,268,623 km^2 (Figure 14) (Marine Conservation Institute, 2024). As with the Faroe Islands, Greenland is a self-governing country and a part of the Kingdom of Denmark with a parliamentary democracy led by the Prime minister of Greenland. The constituent country holds the exclusive right to legislate in most areas excluding foreign and security policy, and monetary policy (The Nordic Council, 2024). Unlike Denmark, Greenland is not a part of the EU meaning that the country is similar to the Faroe Islands insofar as it manages its own fisheries and trade agreements with the EU and other countries (The Danish Parliament, 2021).

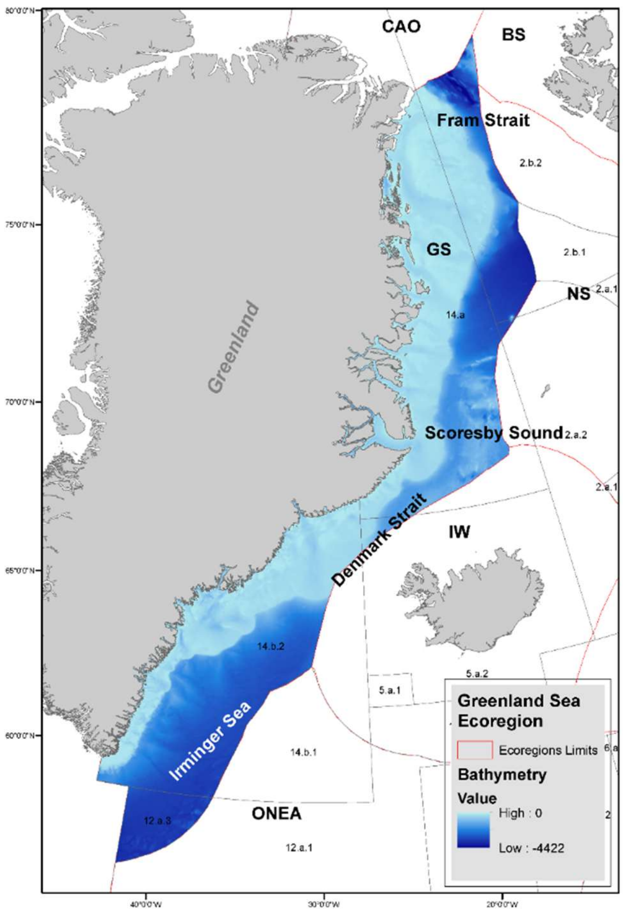

Figure 14. The Greenland Sea ecoregion defined by ICES including other ICES ecoregions indicated with thin grey lines Source: reproduced from (ICES, 2023d).

The Ministry of Fisheries and Hunting is responsible for the management of the fisheries sector in Greenland. Decisions on management, quotas and fishing efforts are based on advice from the Greenlandic Institute of Natural Resources which works closely with ICES (The Ministry of Fisheries and Hunting, 2021). Widely distributed stocks (e.g., herring, capelin and mackerel) that are targeted by the fisheries sector are managed through coastal state agreements and the NEAFC, while Greenlandic halibut and golden redfish are managed through a bilateral agreement between Greenland and Iceland (ICES, 2022a). The fisheries are divided into inshore and offshore fishing (The Ministry of Fishing and Hunting, 2021), and there are two types of fishing licences through which fishing is regulated: fixed-term licence and non-time limited licence (ICES, 2023d).

On 8th January 2021, the EU and Greenland concluded negotiations for a new Sustainable Fisheries Partnership Agreement (SFPA) and a new Protocol that will strengthen their cooperation in the fisheries sector for the next four years. The agreement will allow the EU fleet (12 large-scale industrial trawlers) to continue fishing in Greenland waters for 4+2 years while continuing to contribute to the development of the fisheries sector in Greenland. The EU will provide an annual financial contribution of 16.5 M€ to Greenland based on the fishing opportunities and newly negotiated reference prices. A significant part of this contribution, 2.9 M€ per year, is specifically allocated to promote the development of the fisheries sector in Greenland. The estimated value of the contribution for the entire duration of the protocol is 99 M€. Furthermore, EU ship owners will be required to pay licence fees for fishing opportunities (EU Commission, 2021).

The fisheries sector in Greenland is a crucial industry for the country economy and the second-largest sector by employment. This dependence is particularly pronounced in settlements where fishing and post-harvest employment are the primary sources of income for residents (coastal villages). Approximately 8% are employed in fishing itself (Figure 15) while 16% of the population works in the fishing sector overall (i.e. along the value chain) (Grønlands Statistik, 2023) with approximately one-third of the revenue generated by companies in the country created within the fisheries and fisheries-related industry and trade (5,928 DKK million in 2020). Fish constitute the single most important trading commodity. Two companies, Royal Greenland and Polar Seafood, dominate the land-based fishing industry in Greenland (the former is Greenland's largest company). Recently, a number of private seafood companies have established themselves in the country.

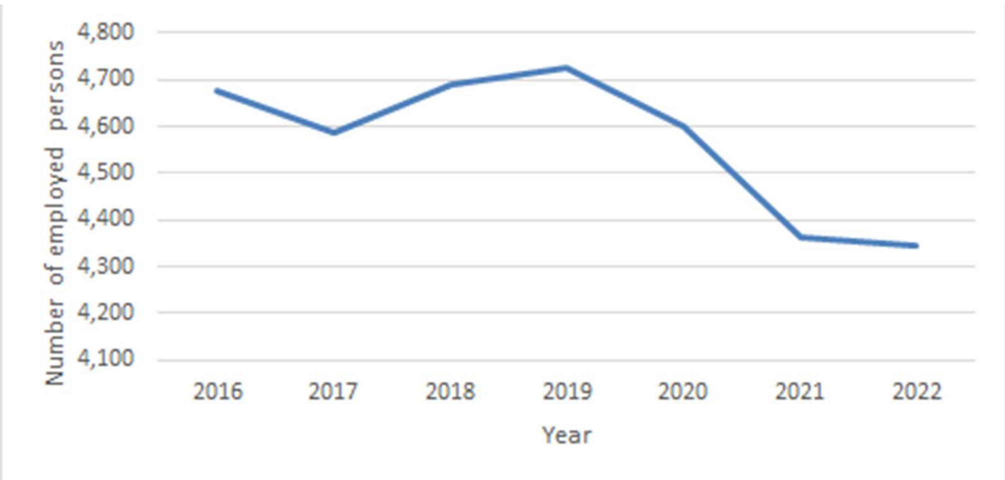

Figure 15. Number of employees in fishing and other related industries in Greenland (Statistics Greenland, 2024)

The entire ecoregion of Greenland is affected by climate change, with the fishing industry being the most important and direct environmental pressure on the ecosystem together with physical seabed disturbance caused by mobile bottom-contacting fish gear (Figure 16). Fishing has impacts on pelagic and benthic habitats and biota while seabed disturbance negatively impacts the benthic habitats and biota (most demersal fisheries are bottom-trawl fisheries) (ICES, 2023d).

In recent decades, sea ice coverage in Greenland has been declining, with a decrease in winter maximum sea ice extent since 1979 and a weak decline in summer minimum ice coverage since 2006. There is evidence of changing surface water temperature and salinity throughout the ecoregion. Surface water temperature has increased by 1-2°C in the narrow southeastern Greenland shelf and in the northern subregion, while it has cooled in the southeastern part of the ecoregion. Surface water salinity has been increasing in open waters of the ecoregion and decreasing in the East Greenland shelf waters and Irminger Sea (ICES, 2023d).

Shifts in the distribution of several marine species have been noted due to environmental pressures:

- Over the last decade, immigration of bluefin tuna (*Thunnus thynnus*), mackerel, and Norwegian spring-spawning herring during the summer (June–September);
- Starting in the 1990s, there has been a westward shift of capelin nursery and feeding grounds;
- Retreating sea ice has increased sightings of whale species, such as fin whales (*Balaenoptera physalus*) and humpback whales (*Megaptera novaeangliae*) on the shelf areas.

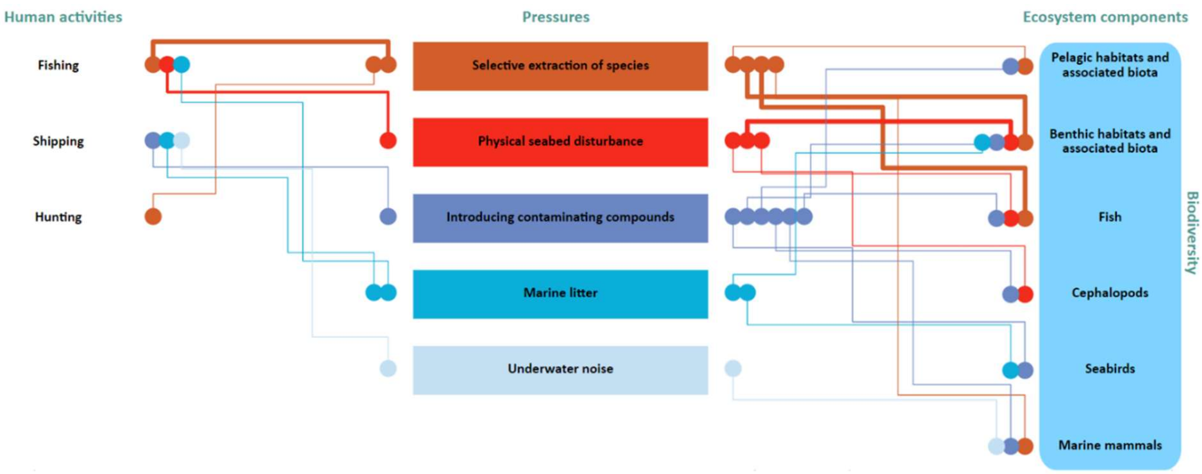

Figure 16. Major human activities, regional pressures, and ecosystem components affected for Greenland. The top linkage chains are responsible for 89% of the risk score in the ecoregion and illustrated as solid lines. The width of lines indicates the relative importance of individual links. Human activities and pressures are listed in decreasing order of their relative contribution to total risk. Source: reproduced from (ICES, 2023d).

Iceland (description, economy and environmental pressures)

Iceland is a volcanic island with a land area of 103,492 km^2 and 6,088 km of coastline where most of the landscape consists of glaciers and big lava fields, and a population of 387,000 people (the largest among countries in this DA) (The Nordic Council, 2024). The Exclusive Economic Zone (EEZ) (Figure 17) has a sea area of 758,000 km^2 (Statistics Iceland, 2024). Iceland is a republic with an elected president who leads the Icelandic parliament. The country is not a member of the EU, but part of the European Economic Area (EEA).

Fishing is one of Iceland's most important sectors and is managed by The Ministry of Food, Agriculture, and Fisheries. However, some of the shared stocks within the Icelandic ecoregion are managed by NEAFC or through coastal state agreements (ICES, 2022a). The country has a long fishing history and was one of the world's front runners in adopting Individual Transferable Quotas (ITQs) in 1979 and has since expanded from including only herring to most fished species (Nordisk Ministerråd et al., 2018). All vessels operate under a TAC system, and fisheries advice is provided by ICES and the Marine and Freshwater Institute of Iceland (MFRI) (ICES, 2022a).

While the Icelandic economy is largely service-based, particularly tourism-related services (the service sector accounts for ~75% of the Gross Domestic Product (GDP)), the fisheries sector is still very important to the economy and culture. Most of the fishing is done by Icelandic vessels and only a small part by neighbouring countries via bilateral agreements. The fisheries sector directly employs approximately 7,500 people (~4% of the total workforce) and contributes 8.1% to the national GDP directly, and 25% if indirect effects of the 'ocean cluster' are taken into account (Iceland Responsible Fisheries, 2021). In 2020, marine products exported amounted to 604,000 tonnes with a value of €1.8 billion and made up 43% of the value of exported goods. Within the sector, cod is the most valuable fish stock, accounting for approx. 49% share of total seafood industry exports by volume in 2021-2022 (Iceland Responsible Fisheries, 2021).

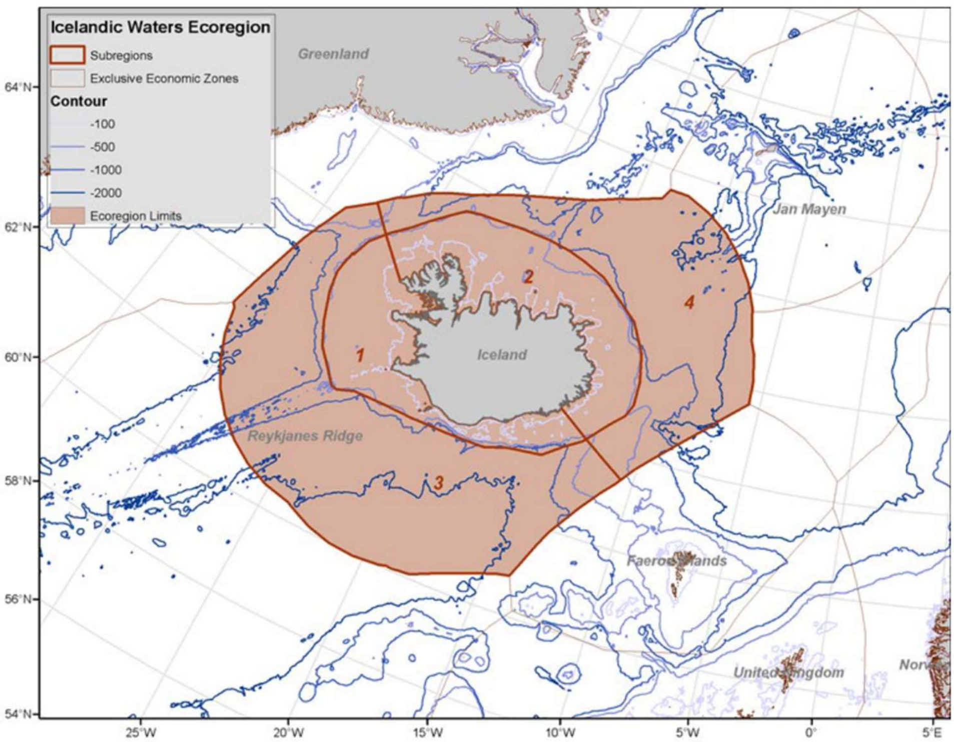

Figure 17. The Icelandic waters ecoregion showing Exclusive Economic Zones, subareas and depth contours (ICES, 2022a).

The main drivers putting pressure on the ecosystem are fishing and coastal development (Figure 18). Fishing is linked to the selective extraction of species (including non-target species), abrasion, and substratum loss and smothering. Abrasion is caused by mobile bottom-fishing gear targeting fish, shrimp and Norway lobster (ICES, 2022a). Bottom trawling has been observed to impact biogenic habitats such as sponge aggregations, coral gardens, and cold-water coral reefs. Other less common abrasion and habitat loss pressures include laying telecommunication and power cables on the ocean floor, anchoring, and static gears. The activity causing a disproportionate amount of sediment smothering is commercial bottom fishing, although lack of data prevents an understanding of the magnitude of the impact (suspected to have decreased in line with reduced trawl fishing activity). In in-shore areas, harbour dredging, aggregate extraction, sediment dumping, cable and pipe laying, and coastal development related to aquaculture and land reclamation have contributed to sediment smothering (ICES, 2022a). Coastal development such as land claim for coastal defence, road building, harbour construction, aggregate extraction and construction of bridges across fjords has contributed to coastal habitat loss.

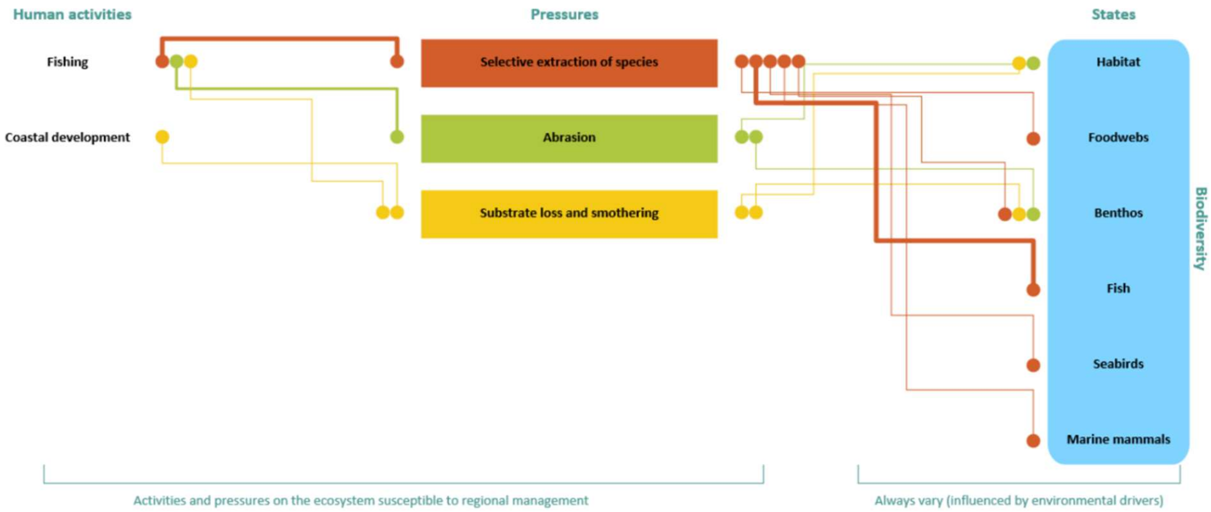

Figure 18. Icelandic Waters ecoregion overview with the major regional pressures, human activities, and state of the ecosystem components. The width of lines indicates the relative importance of individual links (the scaled strength of pressures should be understood as a relevant strength between the human activities listed and not as an assessment of the actual pressure on the ecosystem) Source: reproduced from (ICES, 2022a).

The focus of this study

The focus in this DA is on the pelagic ecosystem of the Northeast Atlantic, namely, the commercial pelagic fisheries as a main anthropogenic driver of ecological (marine ecosystem) impacts across the three countries. The commercial pelagic fisheries sector is a vital part of the economy for each of these countries. The economic importance and overall value of the sector extends beyond the fishing activities and is spread across the entire value chain (local communities, processing/export, small businesses/startups, etc.). As such, the fisheries sector also ties into the social system shaping, forming, and connecting to identities (e.g. "I am a fisher"), as a sense of place/belonging (e.g. "I am part of the fishing community") and cultural/local community/indigenous history. Therefore, altogether, the DA represents a SES with connections within and across geographies, ecosystems, and communities.

In this region, modern pelagic fishing operations are highly industrialized and efficient, conducted by large sea-going vessels that often spend several weeks at sea. Processing plants are often found in smaller towns and may serve as a large source of full-time or seasonal employment. The pelagic species most commonly targeted (mackerel, herring, blue whiting, and capelin) have different life histories and different impacts on food webs, but in common they all are migratory stocks that are fished at various times and locations along migratory routes. As such, they are stocks shared among the Arctic DA countries and the other coastal states (e.g., EU, UK, Norway, and Russia). Many of the large industrial companies fishing the resource are highly vertically integrated (linking fishing, processing, and exporting), and are additionally multinational companies with branches or ownership spanning countries within the DA. As a result, impacts are also linked to geopolitics in the formation or removal of multi-lateral sharing agreements.

In addition to the ecosystems described above by country, a large proportion of the fishing activities for these stocks take place in the Norwegian Sea (ICES, 2022b, Figures 19 and 20). In focusing on pelagic portions of all four ecosystems described, it is of note that nearshore activities have little known impact on pelagic species, and therefore the only human activities deemed consistently important are those that overlap in space with the migratory routes of these species; that is, the fishing activity itself and shipping. As these are some of the most remote industrial human activities in the world, data that specifically reflect the pelagic ecosystem can be quite limited. Furthermore, ecosystem linkages with these highly migratory species are further muddled by early life histories that take place outside the North Atlantic altogether. As a result, few linkages have yet been observed between ecosystem variables and pelagic stock dynamics (ICES, 2023b). Capelin is the only commercial pelagic species of the four known to have strong ecosystem impacts as a food source for important demersal species, such as cod, as well as seabirds and whales.

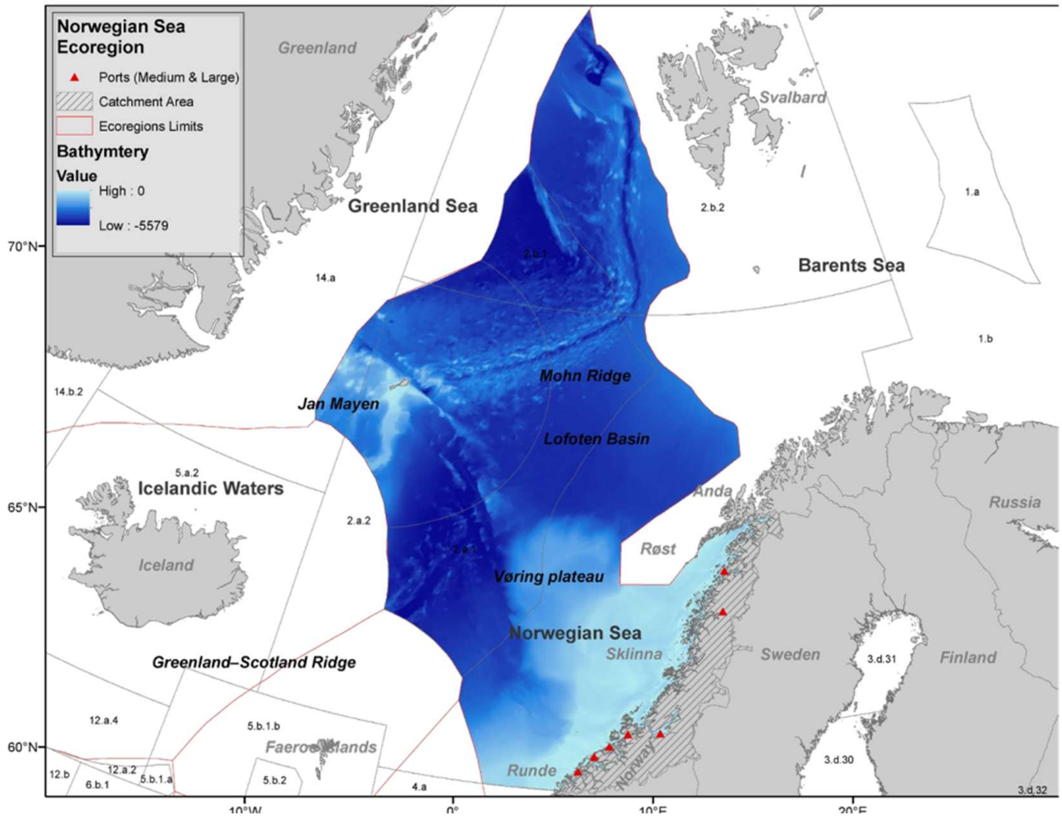

Figure 19. The Norwegian Sea ecoregion defined by ICES including other ICES ecoregions indicated with thin grey lines Source: reproduced from (ICES, 2022b).

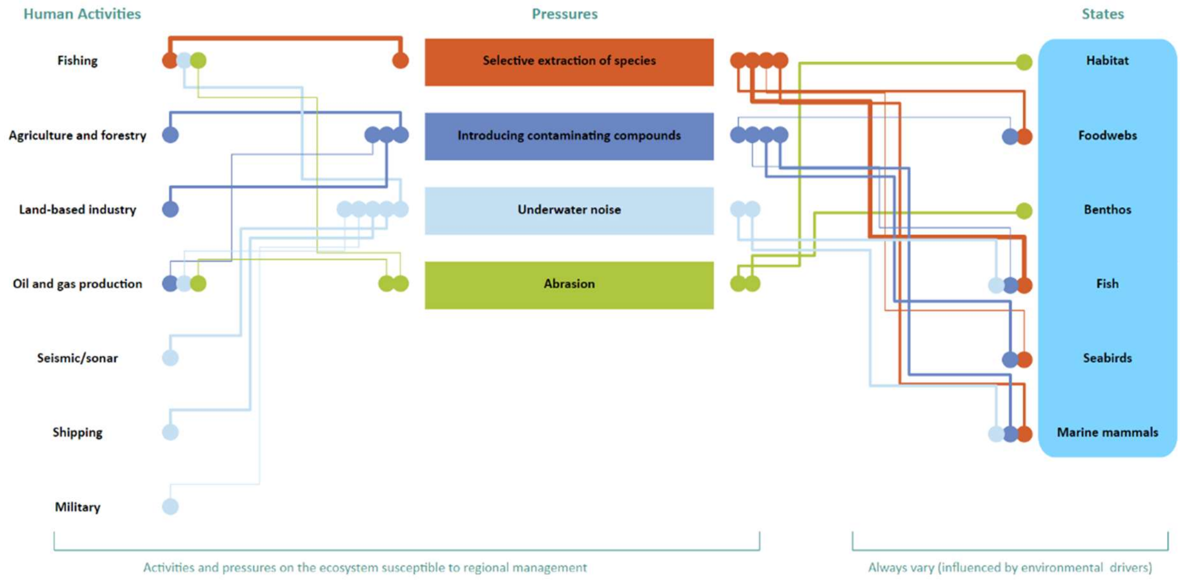

Figure 20. Norwegian Waters ecoregion overview with the major regional pressures, human activities, and state of the ecosystem components. The width of lines indicates the relative importance of individual links (the scaled strength of pressures should be understood as a relevant strength between the human activities listed and not as an assessment of the actual pressure on the ecosystem) Source: reproduced from ICES (2022b).

4 Simple Social-Ecological System model description

4.1 The simple SES (sSES)

The implementation of the sSES approach is based on the Integrated Systems Analysis (ISA) framework (Elliott et al., 2020), which was selected following a comprehensive literature review and an analysis of the strengths, weaknesses, opportunities, and threats (SWOT) associated with existing SES analysis frameworks to assess which of these would meet goals of successful EBM the best (See D3.1: Simple SES literature review (Smith et al., 2023)). The discipline of Systems Thinking provides tools to look at how we manage the marine environment in a rounded and deliberate way. Under this systems approach, we recognise that marine systems are fluid in nature and highly interconnected; hence, embracing this complexity in a structured way can help design programmes of measures for the prevailing circumstances. The objective of this approach is to safeguard the examined SES to sustain the functioning of the ecosystem and hence the provision of valuable services, such as habitats, ecological components and healthy ecosystems that supply the fish and cultural features from which we derive goods and benefits. The sSES is designed to enable an understanding of the dynamic relationships between human communities, stakeholders, institutions, and marine ecosystems. Its primary objective is to guide effective management and policy decisions that align with ecological sustainability and societal well-being.

The sSES approach includes problem-solving methods, qualitative mapping tools and process management resources to be operational. The problem-solving method that underpins the analysis is the DAPSI(W)R(M) method (pronounced *dap-see-worm*) and is a tool to structure issues and priority focuses of an area that affects both people and nature (the social and ecological elements within a system). This framework provides a structured approach for ecosystem-based management by categorising key features of Drivers, Activities, Pressures, State changes, Impacts (on human Welfare), and Response (using management Measures) (Atkins et al., 2011; Elliott and O'Higgins, 2020). The sSES approach provides a method to understand better how different parts of a problem are connected and influence each other when governing marine systems through the use of this structure.

The DAPSI(W)R(M) framework use in the sSES process is multifaceted. Firstly, it provides a holistic lens through which the intricate interplay of socio-economic forces (Drivers) and human actions (Activities) can be linked to their direct effects (Pressures) on natural and societal environmental states. This progression naturally leads to an assessment of how these altered states impact human welfare, thereby closing the loop with Responses that aim to accommodate, mitigate or adapt to these drivers, activities or pressures thereby preventing the effects on the natural and social systems. An aspect of the DAPSI(W)R(M) framework in the context of sSES is its ability to aid in storytelling. This narrative approach is crucial for effectively communicating complex environmental issues to a broad audience, including stakeholders, policymakers, and the general public. By structurally mapping out the sequence from the underlying drivers of environmental change to the societal responses, the framework creates a coherent and comprehensive narrative. This narrative not only enhances understanding but also engages stakeholders in a way that isolated data points or disjointed facts cannot.

Bringing together the DAPSIW)R(M) model, systems thinking in the form of causal loop diagramming and other tools, produces a system for good project and information management; the sSES, therefore, represents a significant shift in our thinking about how we tackle complexity in the marine environment on an ongoing basis as an action learning cycle. Hence, altogether, the process management (The Process and Information Management System (PIMS)), the problem structuring method (The DAPSI(W)R(M)), and the iterative learning cycle compose the sSES approach represented in Figure 21.

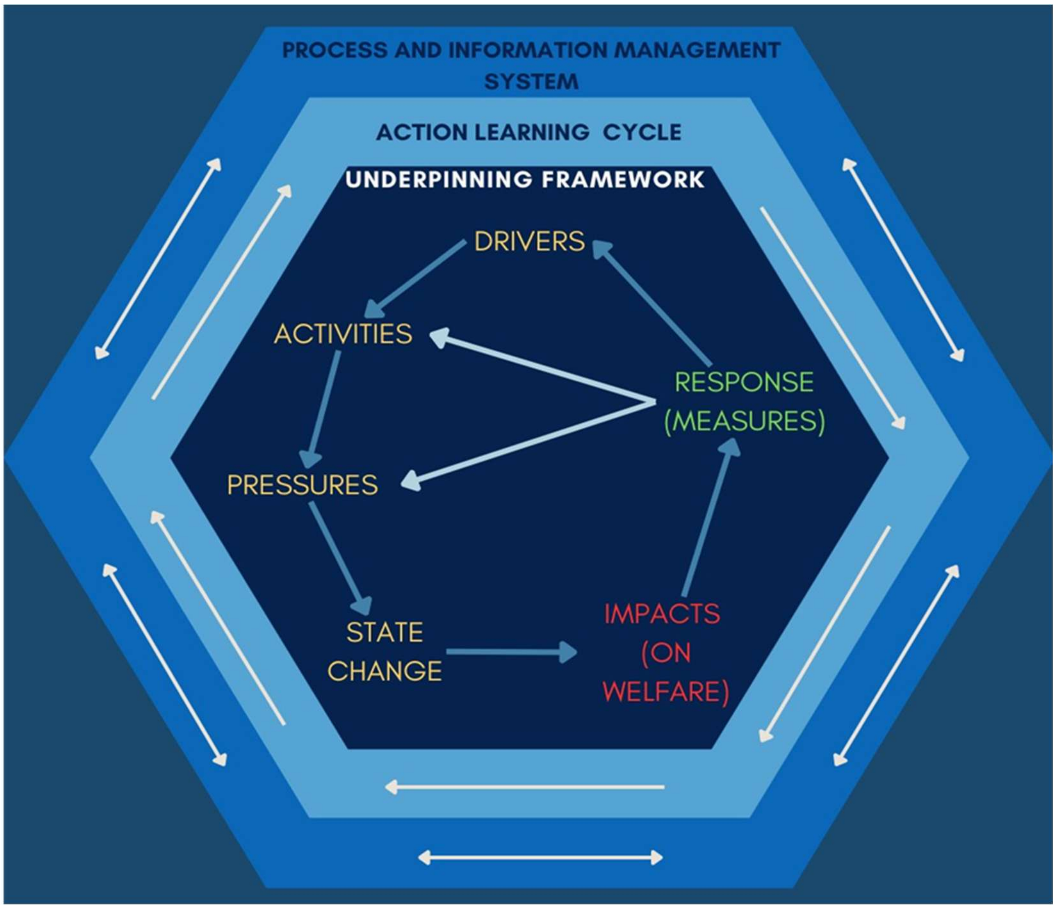

Figure 21. The operationalised Integrated Systems Analysis used in the Simple Social-Ecological Systems approach.

An in-depth summary of the sSES can provide an understanding of the foundational concepts of approach within the Marine SABRES context; this can be found on Page 8 of the Simple SES guidance document (Deliverable 3.1: Gregory et al., 2023). The sSES consists of three parts – A - Setting Priorities, B - Getting the |Information, and C - Using the Information.

4.2 Part A: Setting Priorities

To manage the process effectively, the sSES highlights the importance of management and logistic aspects through the establishment of a Process and Information Management System (PIMS). The PIMS prioritises resource management, stakeholder engagement, communication, data management, evaluation, and governance. Establishing a clear and well-prepared foundation is crucial for performing a successful SES analysis. This includes ensuring good data provenance, assuring the process and its outcomes, and organising social system information (such as legislation, administration, stakeholders, and communication aspects) in a meaningful manner. Another important aspect is setting clear goals and priorities for the process, which helps establish a boundary for the analysis process from the outset; this ensures that adequate resources are available to undertake the scope of the analysis. Hence, the PIMS is a preliminary approach to establish a solid basis for the sSES analysis (See pages 9-23 of the sSES guidance document (Deliverable 3.1: Gregory et al., 2023)).

The initial phase of the sSES methodology involves defining the system scope. This step encompasses identifying the specific marine ecosystem in focus and the corresponding human systems that interact with this ecosystem; the delineation of system boundaries is essential to ensure a focused and relevant analysis. The problem structuring framework requires that the system should be analysed in key components, including resources, actors, governance structures, and external drivers. In a marine context, these resources might encompass biological entities such as fish populations and physical elements such as water quality. At the same time, actors could range from local communities to international regulatory entities.

Within the activities of the Marine SABRES project, the focus issue for each DA has been defined by stakeholder engagement in WP2 (Hummel et al., 2023). In future use, this initial stage will include a process of issue identification through engagement with stakeholders and various tools to do so, for example, rich pictures (See: Bell and Morse, 2013). Following the establishment of a focal issue, the use of stakeholders to identify impacts on human well-being as it may be that we need to prioritise these by, for example, using the Delphi approach (See: Mukherjee et al., 2015). In each DA or component part of a DA in Marine SABRES, the aim is to create a CLD for each of the priority Impacts together with a composite CLD of the Impacts that are focussed on the Issue and over a relevant time-horizon. The creation of one or more CLDs aims to increase our understanding of the behaviour of the system that is causing concern.

4.3 Part B: Getting the information

Using Indicators (See Briefing paper 6: (Atkins and Smith, 2023)) for each element of the DAPSI(W)R(M) we assess each Impact established in Part A and, in turn, identify the relevant measure of the variables and record a time horizon over which the effects on societal goods/benefits are manifested (a time series of the element's behaviour over time). This process is repeated by establishing the various Drivers, Activities, Pressures, and State Changes, their relevant indicators (how to measure the elements), and the behaviour over time (i.e. time series) in relation to the Impact of concern. Using the DAPSI(W)R(M) framework in this way should help set a manageable boundary for the CLD in terms of what is deemed relevant to include and also help to achieve a consistent level of understanding of the spatial and temporal aspects of the SES. To define the Impact dynamically, a review of historical data on key elements and the construction of Behaviour-Over-Time charts for the elements is instructed within the guidance. These charts can serve as reference points throughout the theory-building process, helping focus the conceptualisation and validate emerging theory. This is accomplished when using expert opinion to assess the variables within the adjacency matrices (matrices that assign a positive (+) or negative (-) and the strength of this relation to each combination of the variables in the linkage framework). For detailed Step-By-Step instructions relating to gathering the information for the analysis, see Exercises 1 – 6 in the SES guidance document (Gregory et al., 2023. Pages 42-50).

4.4 Part C: Using the information

The organised method of data collection established by the PIMS section is used to create impact-based CLDs in the Kumu software[1], which illustrate the relationship between the social and ecological elements of a system and their influence on each other. Exploring the behaviour of the system through tracing loops and identifying key leverage points (points within the system where a small shift in one thing can produce significant changes throughout the system) can allow for a meaningful understanding of the SES. Hence, causal theories can be developed to draw out the interrelated behaviour of elements over time, and these theories can be validated by analysing relevant data collected in Part B of the approach. Once the diagrams have been developed and validated (by stakeholders), they can be added to a composite issue-based CLD to gain a better understanding of the system behaviour regarding the focal issue of concern. The behaviour of the system can then be explored through storytelling (See: Vigliano Relva and Jung, 2021) and 'what-if' analysis to identify potential scenarios and their implications on other elements of the system.

4.5 Reflection and Adaptation

The sSES approach requires robust stakeholder engagement and the use of participatory methods prompts (See pages 15 - 19 of the sSES guidance (Gregory et al., 2023)). This approach ensures the inclusion of diverse perspectives in the analysis, enhancing its comprehensiveness and relevance. Such engagement is instrumental in validating the research findings and ensuring that the analysis encapsulates a broad spectrum of viewpoints. Although throughout the nature of the project, there will be a start and finish date for this application, it is emphasised that this approach is beneficial as an iterative and circular process. The overall aim of sSES is to create Response by management Measures that are accepted by decision-makers and implemented in the form of amendments to existing or newly required policies, laws, etc., with the overall aim of achieving the required objectives and vision. Within the PIMS section, there are outcome and process evaluation prompts which promote this reflection of the sSES approach to evaluate if the object has been achieved, how stakeholders were engaged, and lessons learnt for future actions/ processes (See page 21 of the sSES guidance (Gregory et al., 2023)).

[1] Kumu software is a free online tool to create the qualitative systems tool of Causal Loop Diagrams. Link: https://kumu.io/

5 Validation and Testing of the sSES approach

This section covers the theoretical basis that underpins the validation protocol. A detailed review of the field of system dynamics is presented in Appendix 1, where CLDs belong, shows the historical development of how the validation was happening previously, focusing on qualitative models instead of numerical simulations, which is the most appropriate to the present case. In addition, it presents the discussion of semantics regarding validation and justifies the reason for the name "presumed utility" for the validation protocol.

In **this section, we show the short version of the theoretical foundation from which a validation protocol was built** (Table 1) thereby giving a brief version (Appendix 1) of the literature of system science responsible for the criteria used in Table 1.

5.1 Introduction

Two types of models form the main tools in system dynamics: quantitative stock-and-flow models and qualitative Causal Loop Diagrams (CLDs) (Sterman, 2000). The present piece focuses on the CLDs due to their relatively reduced level of complexity and coherence with the qualitative nature of the model we focus on to work, namely, the product of the sSES. In addition, these system dynamics models represent a set of causality models not familiar to most people, which makes this section timely.

Causality models represent a form of a theory of how a system works. It describes the connections of the elements of the system in such a way that one can create an understanding of the system by understanding the causalities described in the model. On the other hand, statistical models are based on ideas of correlation between variables in the system that can be used to forecast or predict the behaviour of a system, preferably inside the same parameter range to which the correlation was observed. However, they cannot produce the same explanation offered by a causal model (Barlas and Carpenter, 1990). Appendix 1 gives a detailed description of the fundamental theory.

This discussion is highly relevant to the validation of system dynamics models and will show a pure statistical validation process, if restricted to a mathematical formal test, which is far from delivering the desired comprehensiveness of a quality enhancement process toward a useful model. Therefore, if one considers that validation should be done strictly via a formal mathematical process, the result would be Boolean (true or false). On the other hand, when a broader perspective is adopted, which is understood to be the most appropriate to the present case, the validation of the model becomes something of a dialogue, iterative, and a process towards learning and participation. In this perspective, models are not necessarily true or false, but open to the new axis of usefulness, under the limitations of a partial, provisional, and socially accepted validity.

What validation is and what it is not

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 modelling process, which the author defines as to "aid in the design of improved industrial and economic systems", consequently defining the ultimate test of validity as "whether or not better [management] systems result from investigations based on model experimentation". However, this ultimate validation test might be far from the modelling process, which calls for an intermediate step of evaluation, closer to the model-building process, that helps to substantiate some confidence in it.

Forrester and Senge (1980) consider that validation is the process of "establishing confidence in the soundness and usefulness of a model as a policy tool". To that end, the confidence in the model must be transferred from the modeler to the users, a step without which, the potential of the model to enhance the management system will not be realized, and thus the model is useless (and invalid). Later on, Oreskes et al. (1994) claimed that verifying or validating numerical models of natural systems is impossible. That happens for two reasons: first, as systems are open, it is implicit that there are variabilities in the system that necessarily were not captured by the model; second, for some results, the more verisimilar they appear it 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. Verifying, thus, can only happen in closed systems when all data are known and known to be correct. This principle is also relevant for qualitative models, therefore the inclusion on this topic.

Finally, the present evaluation protocol adopts a terminology to name an incomplete task, as it is dedicated to evaluating something that is not fully deployed, still provisional. As Meadows (1980) considered the term "utility" more appropriate for an indicator of the quality of a model, here we adopt the term "presumed utility" as the indicator of the quality of CLDs. We did not change the terms used by other authors, so validity is still appropriate but here we will refer to the presumed utility.

5.2 Methods

The Model presumed utility

To create the protocol for understanding/measuring the presumed utility of CLDs (Table 1), we selected the contributions that were relevant from the literature, adapted to our understanding of relevance to qualitative models, complemented with some questions to facilitate the self-assessment grading, and added some tests considered relevant (e.g., learning). The term modelling is interpreted here very broadly, as a process of discussion and formalization (in a qualitative or quantitative model) of an understanding of a system and an issue of interest, made by a few or multiple stakeholders in a determined period. Therefore, these tests (Table 1) are dedicated to increasing confidence in the whole process, not only in the model, assessing in the end the presumed utility discussed above. The user of these guidelines must apply them every time it suits, not only to the final product. We echo the idea that model creation and validation should be done concomitantly, and thus it is highly recommended to use these guidelines at the same time as the model creation.

Table 1. Tests for presumed utility (i.e. validation) in qualitative models

Type	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 modelling 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 modelling 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; 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 represent the behaviour 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 system dynamicist	An affinity with the modeler can enhance positively the modelling process and the Policy Insights or Recommendations (PIoR) implementation. Is it possible to report a positive relationship or atmosphere between the stakeholders and the modelling 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 behaviour, instead of a case-specific behaviour. 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 modelling 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 modelling 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 modelling, during/after the modelling 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 and 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 modelling process concluded within the expectations of time and costs? Are there recommendations to improve the efficiency of the modelling exercise for the next team or exercise?		
	21	Documentation	Refers to the adequacy of the process of making every step in the modelling process replicable by taking a formal process or writing assumptions, discussions, updates, or a change in previous steps regarding the modelling 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 modelling 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, modelling 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 modelling 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 behaviour 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 modelling exercise? Are these changes congruent with the desired state modeled?		

The procedure required to apply these 26 tests is to follow the criteria of Table 1 individually, providing clear answers to each of the questions (Comments column). By reflecting on these comments then, iteratively, trying to improve the approach as it is being produced. These iterative improvements should be briefly described in the Comments column.

In the end, modelers should attribute a grade of trust to each of the tests, using a self-report Likert scale (Jebb et al., 2021). For each of these tests, the modeler should provide a grade from zero to five representing the level of satisfaction with each item of the modelling process (0 - item do not apply, 1 – very dissatisfied, 2 – moderately dissatisfied, 3 – nor satisfied neither dissatisfied, 4 - moderately satisfied, 5 – very satisfied) (adapted from Clark and Watson, 2019). The modeler can iteratively navigate the test protocol to reevaluate each test after some action or improvement in the modelling process is implemented. The final test evaluation should then have a set of grades considered satisfactory by the modeler, stakeholders, and users of the results of the exercise.

The tests were divided into four groups (Broad guidelines and processes, Specific model tests, Policy insights, and reviews, administrative and overview) and to facilitate the discussion, the user must understand the tests and apply them as early as possible considered relevant to the process.

6 Application and testing of the sSES approach

Two steps were taken as part of Task 4.1. The first step refers to the facilitated application of the sSES in each DA, for producing outcomes (i.e. CLDs). The second step refers to the assessment of the level of satisfaction of the users of the sSES (the DA representatives and partners of the project) on both the sSES process and its outcomes, using the evaluation protocol described in Section 5.2.

The application of the sSES was facilitated by a series of on-line workshops and sessions, between the developers of the sSES and the DAs as summarized in Table 2. These sessions were proposed to help the users acquire the new technology and to solve troubleshooting or other issues they had with the terminology and with the system approach underpinning the sSES in general. Constant communication throughout the application process was ongoing via a Q&A shared document and direct communication via e-mails and calls.

Table 2. Simple Social-Ecological System (SES) workshops and sessions organized with the Demonstration Areas (DA) and consortium.

Workshop/ Session Title	Date	Comments
WP3 - Task 3.2 Simple SES Guidance Notes	29/06/2023 (2 hrs)	Three weeks after disseminating the SES guidance, an overview of the approach was presented to the consortium and DAs in the online interim meeting and questions were answered relating to the approach.
WP3 - Task 3.2 Simple SES Guidance Notes	04/07/2023 (3 hrs)	An in-depth three-hour workshop on the approach covering the guidance steps, the Excel sheets and the use of Kumu.
WP3 and WP4 Workshop Simple SES application in Kumu	11/09/2023 (2 hrs)	An online tutorial for using the Kumu software alongside the Simple SES approach. This included a logistical overview of inputting the Excel sheets into the software, as well as identifying loops in the qualitative analysis, and using the presentation features within Kumu's capacity for communicating with stakeholders.
WP4 Q&A	20/09/2023 (1 hr)	Q&A session with the Arctic DA on the Simple SES approach.
Workshop on SES application in the Demonstration Areas	04/10/2023 (3 hrs)	A three-hour hands-on, in-person workshop at the GA of the project where DAs and consortium partners were present. This focused on Part B (Getting the information) and Part C (using the information) of the Simple SES Approach. A detailed overview of the DAPSI(W)R(M) and how to assign indicators to the elements were presented and a supporting infographic was distributed as a quick guide to aid the workshop.
Macaronesia Simple SES hands-on session	16/10/2023 (1 hr)	Q&A session with the Macaronesian DA on the Simple SES approach.
Macaronesia DA follow up on Simple SES	27/10/2023 (1 hr)	Following the last Q&A session, this session answered any follow-up questions relating to the Simple SES approach and its application in the Macaronesian DA.

Arctic DA follow up on Kumu	16/01/2024 (1hr)	This session answered any follow-up questions relating to the Kumu use and some definitions of the systems science for the Arctic DA.
Macaronesia DA follow up on concepts and troubleshooting	01/02/2024 (1.5hr)	This session answered follow-up questions relating to the approach in Macaronesia, some technical questions from the team and the Kumu usage.

7 Results from the sSES application

7.1 Tuscany

The study commenced with an alignment between the overarching goals of the Tuscan Archipelago DA within the Marine SABRES project and the results of stakeholder interviews (WP2). The synthesis of these perspectives revealed the focal point of the sSES: the dual role of tourism as both a driver of economic prosperity and a potential stressor on marine ecosystems and local societies. The ultimate management aim is to delineate sustainable pathways that balance a thriving tourism sector with the critical conservation of marine habitats, particularly the ecologically pivotal seagrass beds.

Following the guidance from WP3, we began by identifying the impacts of tourism on the welfare of island inhabitants and proceeded step by step with the identification of ecosystem services, marine processes, pressures, activities, and human needs.

Our System of focus for analysis is the entire Tuscan Archipelago. We acknowledge that the archipelago comprises seven distinct islands, each with its own unique features such as varying levels of protection, accessibility, presence of MPAs, and whether there is a resident population or not. These differences mean that each island can be viewed as a separate Subsystem within the larger system of the archipelago. Despite these differences, we have chosen to maintain an overarching perspective of the Tuscan Archipelago as a single system. This holistic approach allows us to identify elements that are applicable across the entire archipelago while also recognizing that some of these elements may be more or less suitable for individual islands due to their specific characteristics.

Key Element Identification, Indicators, and Data Provenance

Some comprehensive datasets were collated to encapsulate the elements pivotal to the Archipelago's SES (Table 3).

Table 3. Summary of types, elements and indicators related to tourism in the Tuscany Archipelago National Park (TANP). For each of them its Description and Data Source (if available) are included.

Type	Element	Indicator	Description	Data Source
Human needs	Sense of Identity and Belonging	Resident population	Linked to the resident population.	Annual reports of TANP
Human needs	Food Security	Resident population	Associated with the availability of local marine resources, but we used an indirect indicator.	Annual reports of TANP
Human needs	Self-Actualization (Recreation)	Touristic presences	Reflected in the recreational use of the marine environment.	Annual reports of TANP
Human needs	Self-Actualization (Recreation)	Licences for diving accesses	Reflected in the recreational use of the marine environment.	Annual reports of TANP
Activity	Global CO_2 Emissions	Climate Change	Primary cause of climate change.	EU EDGAR (Emissions Database for Global Atmospheric Research)
Activity	Vessel anchorages	Potential anchorages	Proportional to the density of leisure craft vessel in the TANP	EMODNET

Activity	Artisanal Fishing	Artisanal fishing activity	Derived from the density of fishing boats in the TANP	EMODNET
Activity	Vessel movements and maintenance	Vessel density	Reflects direct and indirect impacts on the marine environment	EMODNET
Activity	Recreational Fishing/Harvesting	Biomass of species removed	Reflects direct impacts on the marine fauna	Data gap
Pressures	Climate Change	Cumulative Intensity of Marine Heatwaves	Indicator of climate change.	NOAA
Pressures	Disturbance of species due to human presence	Area of *P. oceanica* habitat disturbed or lost (km²)	Direct disturbance to *P. oceanica* beds	Data Gap
Pressures	Input of Litter and Nutrient Discharge	Median total number of littered items per 100m²	Pollution Impact	ARPAT
Pressures	Extraction Mortality or Injury to Wild Species	Animals killed or injured per activity or per year (number of bycatch of sea turtles)	Direct impact on wild non-target species	ARPAT
Pressures	Input or spread of nonindigenous species	Number of NIS per area	Related to International Marine Traffic and lessepsian species	Expert opinion
Marine Processes and Functions	Primary Production	PREI index	Provides insight into the health of *P. oceanica* meadows	ARPAT
Marine Processes and Functions	Seawater	Turbidity	Affected by terrestrial inputs, nutrient discharge, sediment resuspension	Copernicus
Marine Processes and Functions	Seawater	Sea Surface Temperature	Related to global Climate Change	Copernicus
Marine Processes and Functions	Food Web Dynamics	Mean Trophic Level of Reef Fish Assemblages	A decrease may signal high fishing pressure.	University of Pisa
Ecosystem Services	Seascapes	Posidonia oceanica Meadows	Offers places for recreation and contributes to biodiversity.	TA National Park
Ecosystem Services	Clean Water and Sediments	Quality of Bathing Waters	Key for human health and marine life.	ARPAT (Environmental Protection Agency of Tuscany)
Ecosystem Services	Coastal and Marine Biota	Biomass of reef fish	Essential for a robust ecosystem.	University of Pisa
Ecosystem Services	Carbon Sequestration	Amount of carbon dioxide sequestered	ES provided by *P. oceanica* beds	Data Gap
Ecosystem Services	Natural Hazard Protection	Energy dissipation capacity	ES provided by *P. oceanica* beds	Data Gap
Goods and Benefits	Income and Employment	Employment Rate	Provides income and employment opportunities to the local community.	ISTAT (Italian Institute of Statistics)
Goods and Benefits	Housing	Housing Cost-to-Income Ratio	Represents the balance between living costs and earnings for residents.	ISTAT
Goods and Benefits	Fish for Human Consumption	Fish landings	Integral to the area's culture and tracked through fish landing data.	ISTAT
Goods and Benefits	Use of Places and seascapes	Quality of Bathing Waters	Reflects visual appeal and quality of coastal bathing areas, important for tourism and health.	ARPAT (Environmental Protection Agency of Tuscany)
Goods and Benefits	Aesthetic Benefit	Natural habitats	Crucial for tourism and local well-being, but quantification is challenging.	Data Gap

The modelling of the sSES for the Tuscan Archipelago encountered various data challenges. One of the primary issues has been the lack of long-term, high-resolution datasets that are tailored specifically to the conditions of the Archipelago. For instance, while provincial-level socio-economic data from ISTAT (Instituto Nazionale di Statistica), inclusive of mainland municipalities, provides valuable insights, it lacks the fine detail required given the unique contexts of the islands. Furthermore, the current practice of data collection and reporting is characterized by fragmentation and is often conducted by multiple agencies. This fragmentation makes it difficult to integrate datasets into a consistent long-term series. Compensating for the absence of certain direct indicators required relying on indirect indicators, as long as these proxies could offer us relevant insights. For instance, the density of leisure crafts has been used as an indirect measure of anchoring activity. Nevertheless, there have been cases where data collection and the reconstruction of temporal trends were not feasible. Overall, there is a noticeable heterogeneity in the quality, timing, and spatial resolution of the collected data, which, in some cases, has led to a lack of temporal overlap between datasets for various elements.

In tackling the data challenges identified during the SES analysis of the Tuscan Archipelago, we recognize an invaluable opportunity to advance our understanding of this intricate social-ecological system. The holistic perspective enforced by the SES framework has underscored the need for a more refined and comprehensive data collection approach.

Causal Loop Diagram

In developing the CLD, the guidance suggests utilizing temporal trends in indicators (Behaviour Over Time -BOT-) to infer causal relationships between elements. However, due to the pitfalls of inferring causality from temporal correlations alone, we integrated expert judgment and a thorough review of scientific literature to reconstruct cause-effect relationships within the SES. Through this approach, we derived the strength and direction of interactions, recognizing that many relationships within the SES might be more complex and/or context-dependent.

It was challenging to find clear temporal correlations that aligned with our deduced cause-effect relationships. Consequently, the presumed causal relationships often did not correspond directly with the observed temporal trends among indicators, reflecting the difficulty in establishing definitive links within such a diverse and multifaceted system.

The complete CLD for the Tuscan Archipelago was composed of 26 elements and 53 links (Figure 22). The original number of identified elements was 28, but before importing data to the Kumu software we combined the needs of island inhabitants into one element and joined the ecosystem services provided by *P. oceanica* beds.

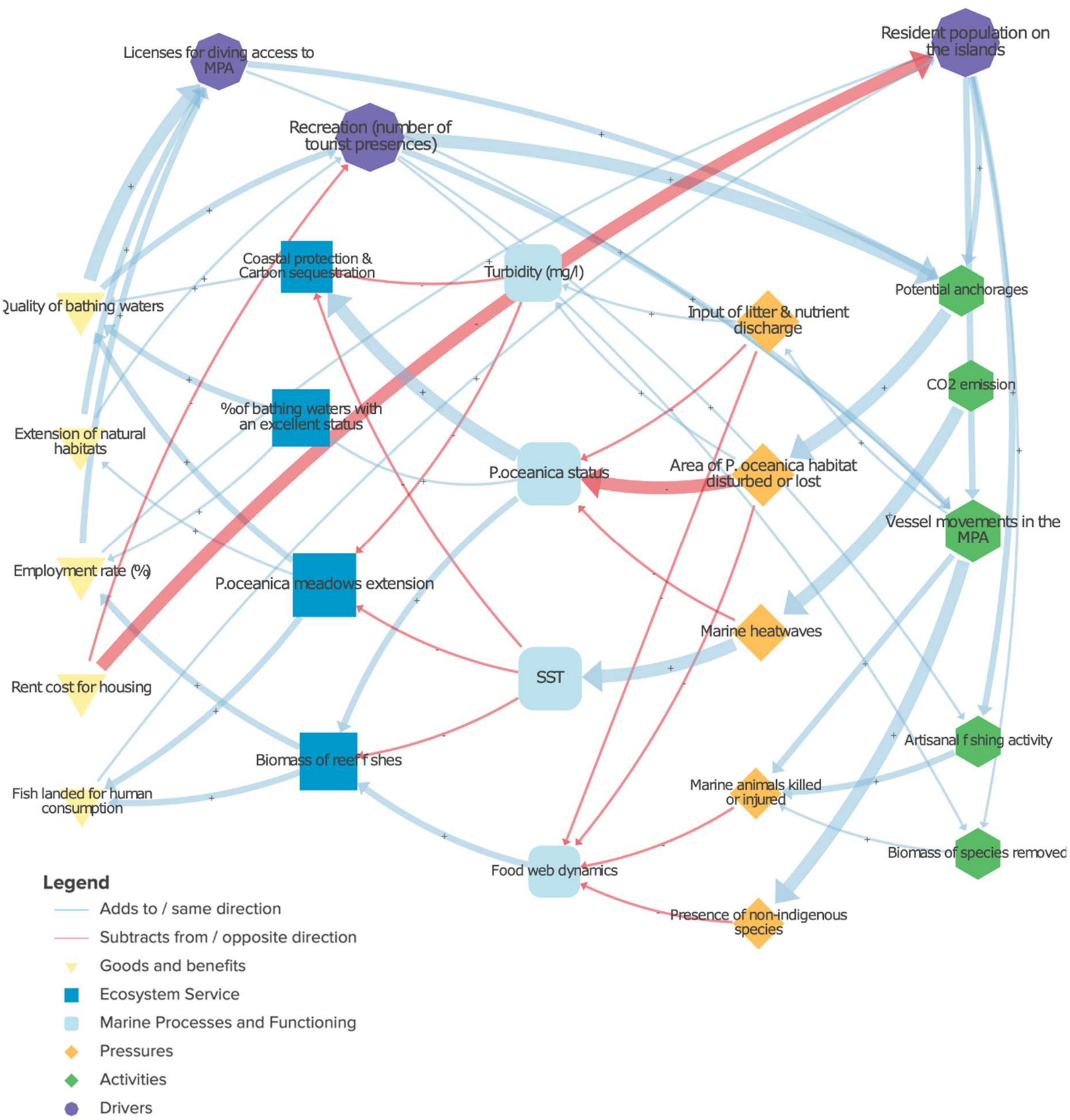

Figure 22. Causal Loop Diagram representing the Tuscan Archipelago Social-Ecological System. SST: Sea Surface Temperature.

Following the SES guidance document, a simplification process was undertaken, which involved combining certain elements to streamline the representation of the system and the removal of exogenous elements (Figure 23). For instance, elements and paths related to climate change were eliminated, as the causes of global CO_2 emissions on coastal ecosystems and related increase of Sea Surface Temperature (SST) and marine heatwaves are significant but not manageable at a local level. Similarly, the housing cost-to-income ratio was excluded from the CLD as it reflected an exogenous element being mainly influenced by broader national instead of local economic factors.

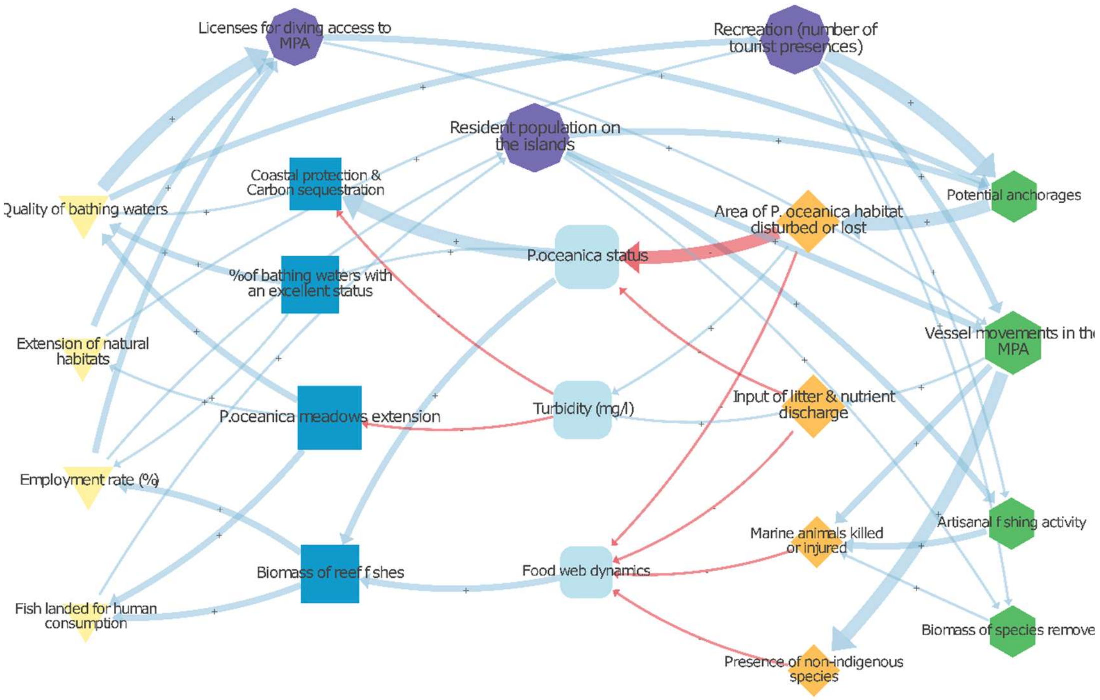

Figure 23. Refined Causal Loop Diagram representing the simplified Tuscan Archipelago Social-Ecological System.

After the simplification process, the network was refined to 22 elements and 45 links (Figure 23). A total of 348 loops were identified, demonstrating the feedback mechanisms at play. Analysing all the loops was a long process and identifying the most relevant loops appeared to be a subjective rather than an objective process based on formal analysis.

One illustrative feedback loop identified in our CLD is the 'Tourists-Nutrient Discharge-Habitat Degradation' loop (Figure 24). This negative feedback loop suggests that an increase in tourist numbers can lead to a rise in nutrient discharge into coastal waters, often as a result of increased waste and sewage that are not adequately managed. Excess nutrients can cause habitat degradation. As these habitats deteriorate, the area may become less appealing to tourists, potentially leading to a decline in tourism over time. This effect can be further magnified by climate change, as the holiday season extends beyond traditional periods with global warming.

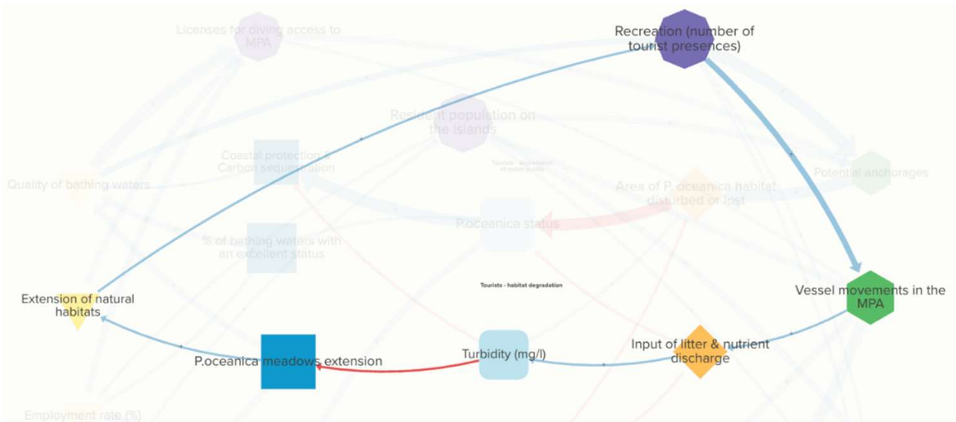

Figure 24. Negative feedback loop indicating the potential deleterious effects of over tourism on habitat quality which in turn may reduce their appeal to tourists.

Furthermore, we conducted an overview of the properties and roles of elements within the CLD, which assists in pinpointing priorities for intervention and management. In Table 4, we have ordered the elements based on their indegree (the number of elements influencing a given element), outdegree (the number of elements influenced by a given element), and betweenness centrality (a measure of the element centrality within the network). Some elements that feature prominently in Table 4 include the vessel movements in the MPA, the number of potential anchorages and the input of litter and nutrients into the sea, and their effects on *P. oceanica.* These elements have shown to be highly influential within the CLD and therefore may represent key targets for management interventions by both the Park Authority and entities responsible for control and monitoring.

Table 4. Overview of the properties and roles of elements within the Causal Loop Diagram, showing the first 10 elements listed on the base of: indegree, outdegree and betweenness centrality. MPA: Marine Protected Area.

	Indegree	Outdegree	Betwenness
1	Food web dynamics	Recreation (number of tourist presences)	Biomass of reef fishes
2	**Vessel movements in the MPA**	Resident population on the islands	**Vessel movements in the MPA**
3	Quality of bathing waters	**Vessel movements in the MPA**	Resident population on the islands
4	Licenses for diving access to MPA	*P. oceanica* status	Food web dynamics
5	Marine animals killed or injured	*P. oceanica* meadows extension	Employment rate (%)
6	**Potential anchorages**	Area of *P. oceanica* habitat disturbed or lost	**Potential anchorages**
7	Recreation (number of tourist presences)	**Input of litter and nutrient discharge**	Area of *P. oceanica* habitat disturbed or lost
8	Resident population on the islands	Quality of bathing waters	**Input of litter and nutrient discharge**
9	*P. oceanica* status	Licenses for diving access to MPA	Licenses for diving access to MPA
10	Employment rate (%)	Employment rate (%)	Turbidity (mg/l)

Causal Loop Diagram Validation

The validation of the CLD for the TA was conducted by both modelers and stakeholders including academic groups as well as regional and national environmental agencies (i.e., ARPAT and ISPRA).

The validation process identified elements of high satisfaction:

- **Purpose:** The objective of the model was easily understood;
- **Meaningfulness of the Process:** Stakeholders appreciated the experience of delving into the model;
- **Engagement:** The participatory approach increased stakeholder engagement and ownership.

Conversely, some aspects received lower satisfaction scores:

- **Replicability:** All stakeholders expressed doubts about the replicability of the results;
- **Extreme Conditions:** Questions regarding the CLD predictive capability under extreme conditions may indicate a gap in the proposed methodology for analysing SES behaviour;
- **Loop Polarity:** Stakeholders struggled to understand loop implications on system dynamics, suggesting the need for more explicit instructions to conduct the analysis;
- **System-Improvements:** It was unclear how this test should be conducted.

During the validation of the CLD for the Tuscan Archipelago, stakeholders indicated areas of both strengths and challenges in the CLD development and analysis. Stakeholders expressed concerns about the transferability of the CLD's dynamics to end-users (e.g. MPA managers). In addition, they observed that while the comprehensive CLD could be overwhelming and difficult to interpret, isolating individual loops might oversimplify the system.

Stakeholders unanimously recognized the inherent complexity of the CLD, particularly highlighting the difficulties in interpreting and analysing a CLD with many loops. Stakeholders also faced difficulties in interpreting and analysing over 300 loops. Identifying the most relevant loops was perceived as a subjective process, that may strongly depend on the expertise and perception of the modeler'. These concerns are consistent with the validation results. Indeed, replicability and the capacity of the model to infer system behaviour under extreme conditions (without a formal procedure in the guidance) were identified as less satisfactory elements.

While stakeholders acknowledged the CLD potential, they pointed out significant barriers to realizing that full potential. The primary issues identified were related to communication challenges and the replicability of CLD results. The feedback highlighted a gap in the guidance for replicating the CLD analysis across different contexts and for inferring system behaviour under extreme conditions without a formal analytical procedure.

Concluding remarks

During the application of the sSES approach for the Tuscan Archipelago we faced various problems, especially the scarcity of adequate data sets. The lack of data, particularly long-term, high-resolution datasets, highlighted a critical need for the creation of targeted data sets for our DA, particularly socio-economic trends and ES.

The SES modelling process, as prescribed by the guidance, should involve an iterative development and validation process. However, the lack of automation and direct linkage between data sets (collected in Excel files) and the Kumu analysis software represented a significant challenge for our analysis. Indeed, adjusting the model with new elements or modifying connections within the CLD required manual updates on both platforms, slowing down the process.

Stakeholder CLD validation showed several areas of high satisfaction, including the purpose, engagement and meaningfulness of the process. On the other hand, replicability, the ability to infer system behaviour under extreme conditions, and loop analysis were the least satisfactory elements that require improvement. The effective communication of CLD results was also considered a barrier to the widespread application of the framework.

In conclusion, our application of the SES framework in the TA may offer suggestions for further improvements. Future SES modelling would benefit from tools that offer automated integration of data updates, allowing for real-time adjustments of the model. A refinement of the guidance, particularly the section dedicated to the CLD analysis, may facilitate the ability to replicate the CLD analysis and its findings across different contexts or modelers/users.

7.2 Macaronesia

The original plan for Macaronesia DA focused on three different aspects of marine conservation: marine protected areas, marine wetlands and ecological corridors. After the interpretation of the results gathered by stakeholder interviews (WP2), the focus of this study was reduced to the relationships between tourism and marine protected areas.

Having the DAPSI(W)R(M) framework in mind, we proceeded by carefully selecting indicators for each level of the analysis (i.e. *Type* in Table 5) that could relate the economic and ecological value of MPAs and the potential impact of human activities (e.g. tourism pressure) on the MPAs themselves.

Table 5. Selected indicators related to tourism inside Marine Protected Areas in Macaronesia.

Type	Element	Indicator	Description	Data Source
Needs	Improved human health benefits and wellbeing	Resident in Macaronesia	# residents in Macaronesia	Census
Activities	Aquatic Sports	Aquatic Sports	# register athletes in: Scuba diving; canoeing; jet skiing; fishing; surfing; sailing; Stand up paddling	DRD
Activities	Recreational Diving	Recreational Diving	# scuba dives that had dove in MPAs each year	CPVP - Azores
Dominant Pressures	Fishing	Fishing	Total Fishing landings (in tons per year)	PORDATA
Dominant Pressures	Input or spread of non-indigenous species	Non-native species	# of non-natives species within years	CIBIO-Açores; MARE-Madeira

Dominant Pressures	Input of litter	Litter recovered per cleaning activity	# organized activities conducted to collect marine litter on the coast	DRPM - DQEM; OSPAR
Dominant Pressures	Disturbance of species due to human presence	Touristic pressure	# tourists inside MPAs	OTA; DREM
Marine Process and Functioning	Food-Web dynamics	Predator/prey ratio of fish within MPAs	# predator species/ # of prey species within MPAs	MARE-Madeira IMAR - Azores
Marine Process and Functioning	Primary production	Productivity	Chlorophyll *a* concentration (annual average)	OOM
Marine Process and Functioning	Formation of species and habitats	Abundance of large predator fish within MPAs	# *Mycteroperca fusca*; *Serranus atricauda*; *Epinephelus marginatus* /m^2	MARE-Madeira IMAR - Azores
Ecosystem Service	Biodiversity	Abundance of benthonic fish within MPAs	# total individuals /m^2	MARE-Madeira IMAR - Azores
Ecosystem Service	Places and seascapes	NATURA reserves	# MPAs (Rede NATURA) each 10 years	DRPM; IFCN
Ecosystem Service	Clean water and sediments	Blue Flag status area	# Blue Flag status areas each year	Bandeira Azul
Good and Benefit	Income and Employment	Marine tourism companies	# marine touristic companies	DRPM and ACIF
Good and Benefit	Food for human consumption	Fish landings of key coastal species	*Mycteroperca fusca*; *Serranus atricauda*; *Epinephelus marginatus* (in tons)	PORDATA
Good and Benefit	Uses of places and seascapes	Area of coastline protected by MPAs	% of coastline protected by MPAs compared to the territorial sea	DRPM; IFCN
Good and Benefit	Education & Research	Research papers concerning MPAs in DA	Number per year (expert view)	Web of Knowledge
Good and Benefit	Education & Research	Number of Blue Schools	Number per year (community view)	ABAE; DRM

Indicator selection, data collection and establishing connections

This DA includes two Autonomous regions of Portugal (Azores and Madeira) each with it is own local government, Research and Development institutions, and different social contexts. These characteristics significantly influenced the selection of indicators since data availability and methodology of sampling differed between the archipelagos.

Data collection and treatment to fill in ISA spreadsheets in a representative way for the DA required considerable time (i.e. months). In almost all indicators, the average value was used since data sources were different. In three cases (i.e., recreational diving, litter activities, and tourism pressure), gathering data from both archipelagos was not possible. Moreover, data coming from research institutes or the government were restricted to specific locations making their scalability less realistic for the entire DA. Finally, for the indicator representing "Clean water and sediment", a proxy was selected to overcome the data availability issue. In addition to these limitations, the identified indicators do not have data gaps and thus the model is as quantitative as possible.

After gathering all the relevant data, the users we proceeded to establish the connections (causal relationships) between the various indicators following the DAPSI(W)R(M) framework. For the most part, they used the collected data to establish whether the relationships were positive or negative. However, in some cases, the outcome did not agree with their expert judgment (Tables 6 to 11).

*Table 6. Links between Ecosystem services and goods and benefits, in Macaronesia. *=Mismatch between BOTs data and expert assessment.*

Macaronesia		Goods and Benefits				
		Tourism companies	Fish landings of key coastal species	Area of coastline protected by MPAs	Research papers concerning MPAs in DA	Number of Blue Schools
Ecosystem Services	Abundance of benthic fish within MPAs	Weak +*	Weak +	Strong+*	Medium +	Weak +
	NATURA reserves	Weak +	Medium +	Strong+	Medium +	Medium +
	Blue Flag status	Medium +	Medium +*	Medium +	Medium +	Medium +

For instance, the indicator Biodiversity (Ecosystem Services), which was defined as the abundance of benthic fish within MPAs, and the indicator Food Web Dynamics (Marine Process and Functioning), which was defined as the predator/prey ratio of fish within MPAs, showed distinct temporal trends (Table 7).

Table 7. Links between marine process and functioning and ecosystem services, in Macaronesia=Mismatch between BOTs data and expert assessment.*

Macaronesia		Ecosystem services		
		Abundance of benthonic fish within MPAs	NATURA 2000 reserves	Blue Flag status
Marine Process and Functioning	Predator/prey of fish ratio within MPAs	Strong + *	Medium+	Medium +
	Productivity	Strong +	Weak +	Weak +
	Abundance of large predator fish within MPAs	Strong +	Strong +	Medium +

While Biodiversity decreased over time, Food Web Dynamics increased over time. Accordingly, there should be a negative causal relationship between the two. This is, however, contrary to our expert judgment and to the wider literature suggesting that MPAs have both a positive effect on the numbers of predatory fish and the overall abundance of fish. In these cases, the users overrode the underlying quantitative metrics with their expert knowledge of the system in the model creation. This discrepancy between what was the expected relationship and that derived from the quantitative data collected may be explained by the fact that not all MPAs are fully enforced which may hamper the establishment of their full potential as a conservation tool (Table 8).

*Table 8. Links between pressures and marine process and functioning, in Macaronesia. *=Mismatch between BOTs data and expert assessment*

Macaronesia		Marine Process and Functioning		
		Predator/prey of fish ratio within MPAs	Productivity	Abundance of large predator fish within MPAs
Pressures	Fishing	Medium -	Medium -	Medium - *
	Non-native species	Medium - *	Weak -	Medium -
	Litter recovered per cleaning-up activity	Medium -	Medium -	Medium -
	Touristic pressure	Weak -	Weak -	Weak -

In other situations, there was also a potential mismatch between what is expected by a particular indicator and what was shown by the available data (Tables 9-11). For instance, productivity is an important bottom-up driver of marine coastal ecosystems. In the oceanic setting, the concentration of ocean *chlorophyll a* was used as a measure of productivity. However, productivity was rather constant over time, making it difficult to establish relationships with other indicators. Moreover, while Touristic Pressure (Pressure) is likely to have a negative impact on some facets of productivity (e.g. secondary productivity), it is unlikely to have a direct causal (or yet unproven) influence on oceanic primary productivity (e.g. *chlorophyl a*). This suggests again that establishing relationships between indicators is not always a straightforward exercise, requiring careful consideration and expert input. This also highlights that the choice of different types of available data (primary vs. secondary productivity) underlying specific indicators may affect the interpretation of the causal relationships.

*Table 9. Links between activities and pressures, in Macaronesia. *=Mismatch between BOTs data and expert assessment.*

Macaronesia		Pressures			
		Fishing	Non-native species	Litter recovered per cleaning-up activity	Touristic pressure
Activities	Aquatic sports	Weak -	Medium +	Weak +	Weak +
	Recreational diving	Weak -	Medium +	Weak + *	Strong +

*Table 10. Links between drivers and activities, in Macaronesia. *=Mismatch between BOTs data and expert assessment.*

Macaronesia		Activities	
		Aquatic sports	Recreational diving
Drivers	Resident	Medium +	Medium +

*Table 11. Links between goods and benefits and drivers, in Macaronesia. *=Mismatch between BOTs data and expert assessment.*

Macaronesia		Drivers
		Resident
Goods and benefits	Tourism companies	Medium +
	Fish landings of key coastal species	Medium +*
	Area of coastline protected by MPAs	Weak +
	Research papers concerning MPAs in DA	Weak +
	Number of Blue Schools	Weak +

Causal Loop Diagram

The complete CLD for the Macaronesia DA is presented in Figure 25. The entire CLD consists of 18 elements and 51 links. Of the latter, 12 were established as negative which included all the links between the four pressures and the three marine processes and functions, as well as two links between activities and the pressure of "fishing". The remaining 39 links were positive.

Compared to the Tuscan DA, which also focused on the relationship between tourism, marine protected areas and conservation, there were some perspective differences when ascribing types of indicators. For instance, in the CLD for Tuscany, recreation (number of tourists) was established as a driver, whereas in Macaronesia's CLD we perceived tourism in two distinct ways: (i) as pressure on marine ecosystems, because an increasing number of tourists exerts increasing pressure on the marine ecosystem as a whole (from, e.g., boat traffic, underwater noise, effluent output); and (ii) as a good and benefit, as increasing tourism creates economic and societal benefits for the resident society.

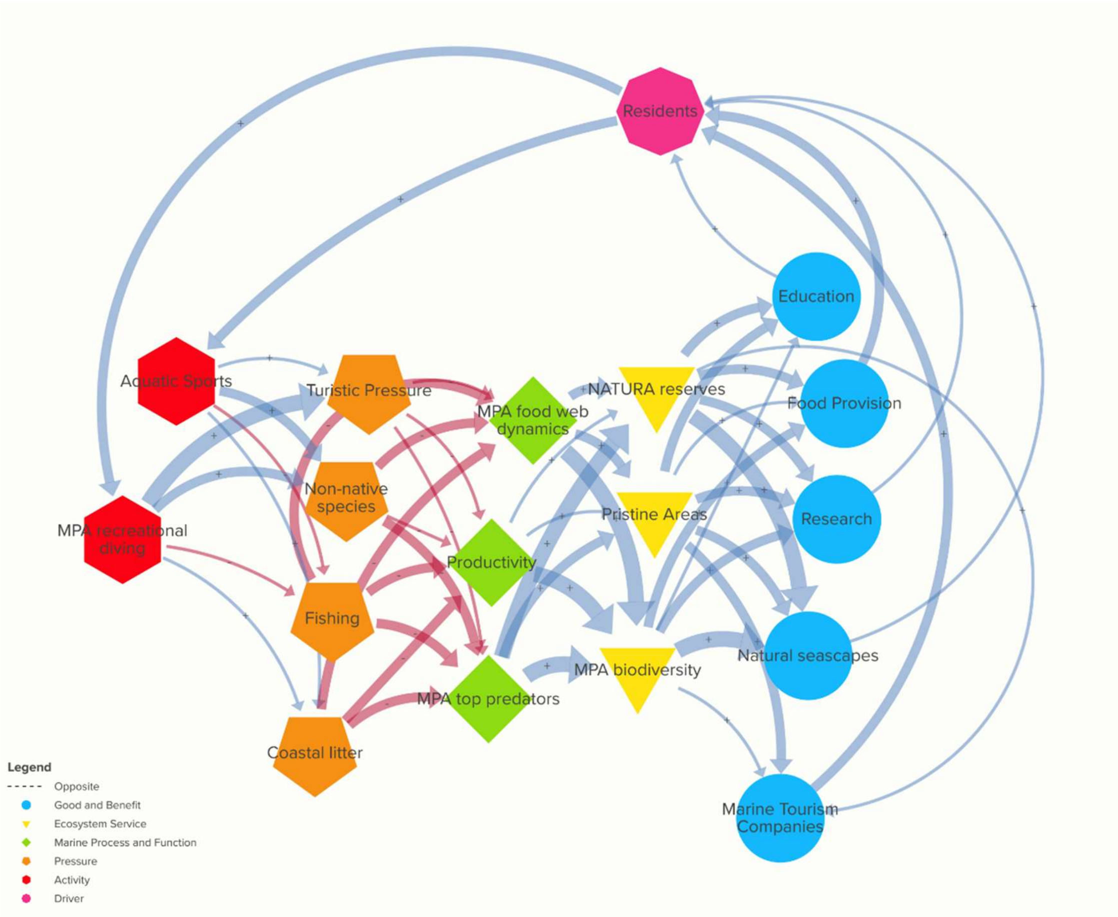

Figure 25. Causal Loop Diagram (CLD) for the Macaronesia region.

Analysis of the number of links between elements in the CLD identified all the Ecosystem Services as the main elements in the model with the highest levels of links (Table 12), suggesting these elements play a central role (possible leverage points) in the modelled system. Moreover, the elements with the widest reach in terms of influencing the system (those with the greater number of outgoing links) were all the Ecosystem Services as well as all the Activities. Whilst Ecosystem Services are not amenable to direct management, Activities are. In contrast, the elements that were most likely to be influenced (in terms of the greater number of incoming links) were Marine Processes and Functions. As with Ecosystem Services, Marine Processes and Functions are unlikely to be directly manageable, but they can be influenced by Activities. This suggests that by managing activities, in our case, recreational diving and aquatic sports, we may have the widest impact on our system via changes in Marine Processes and Functions that underlie the Ecosystem Services we derive from marine ecosystems.

Table 12. Number of links (incoming, outgoing and total) of the various elements in the Causal Loop Diagram of Macaronesia. In bold, the elements with the greatest number of links. MPA: Marine Protected Areas.

Element	Outgoing links	Incoming links	Total links
MPA biodiversity	**5**	3	**8**
NATURA reserves	**5**	3	**8**
Pristine Areas	**5**	3	**8**
MPA food web dynamics	3	**4**	**7**
MPA top predators	3	**4**	**7**
Productivity	3	**4**	**7**
Aquatic Sports	**4**	1	5
MPA recreational diving	**4**	1	5
Coastal litter	3	2	5
Fishing	3	2	5
Non-native species	3	2	5
Residents	2	3	5
Touristic Pressure	3	2	5
Education	1	3	4
Food Provision	1	3	4
Marine Tourism Companies	1	3	4
Natural seascapes	1	3	4
Research	1	3	4

Concluding remarks

During the application of the sSES for the Macaronesia DA, the greatest challenge was represented by the selection of indicators. This difficulty was mainly due to a lack of data or inconsistency in time and spatial scale. Hence, to avoid data gaps, in some cases, it was necessary to rely on proxies, such as the use of the "number of beaches that received the certification of the blue flag" as an indication of the number of coastal areas with good water quality. The relationship between this indicator and the element, however, must be interpreted with caution since the attribution of the "blue flag" does not only rely on the quality of the water but also on other criteria (i.e. environmental education activities, adequate information about local eco-systems, environmental and cultural elements; map indicating different facilities; code of conduct).

The lack of ready-to-use data to fill in the data set (excel file) substantially limited the selection of indicators and consequently the Causal Loop Diagram.

During the application of the sSES approach, it was fundamental to have direct access to the guidance and to the developer of the approach. Several issues arose in the phase of elements and indicators selection, as well as in the final phase regarding the use of Kumu analysis software. The first output of the CLD required adjustments, and 8 connections had to be changed. These modifications were conducted during a DA meeting highlighting the need for expert judgment in the CLD creation. In addition, a stakeholder validation of the CLD would be recommended, to show CLD results, and thus possibly facilitate the application of the framework.

In summary, careful selection of indicators plays a central step in model building. These must be based on trustworthy and reliable sources of data, have temporal and spatial consistency, and be relevant to the metric being examined. Despite the plethora of data currently available online, it is still difficult to find suitable indicators that directly reflect the problem being studied. The choice of indicators used in our model was often based on proxies since direct links between the topic and data availability or reliable sources of data were unavailable in many situations. One additional difficulty was data interchangeability and integration when these are collected under different legal frameworks, especially in situations, such as in our DA, where the model focused on a geographically wide region which is under the jurisdiction of distinct regional policies (e.g. Azores and Madeira Regional Governments).

7.3 Arctic

Using the guidance of the integrated systems approach (ISA) developed by WP3 entailed a careful selection of indicators to represent the DAPSI(W)R(M) features relevant to the Arctic DA and its topic in focus. As noted in section 3.3, the topic in focus in this DA is commercial pelagic fisheries in the Northeast Atlantic. According to the WP3 guidance, the starting point for collecting information for the ISA is unfolding complexity and identifying impacts on welfare as it derives from the presence of commercial pelagic fisheries - Exercise 0 (Gregory et al., 2023, p. 41). The level of consideration is twofold: System in Focus and Sub-systems. Overall, the focus is on the spatial area being studied, the Arctic (System in focus). But in completing the ISA, the focus moved from the Arctic to the smaller specific areas within the Arctic DA, that being the Faroe Islands, Iceland, and Eastern Greenland (Subsystems), to take into account the differing social and economic features of each area.

The timeframe selected by the regions differs. In the Faroe Islands, local indicators were set at 20 years. This timeframe was selected to reflect on the changes, developments, and challenges present in the Faroe Islands regarding their dependence on the marine environment. There were discussions that longer timeframes would be more representative, considering that the marine environment and industrial, societal, and economic elements are developed over longer periods. In Iceland, the timeframe considered applicable was 11-12 years (i.e. from 2011 onwards to the most recent year possible). This was chosen as it reflects the most recent changes in the ecosystem as well as the timeframe considered in the interviews (10 years). In Greenland, the timeframe considered was the same as for Iceland, as far as possible. Finding data for some indicators was difficult in some cases, resulting in shorter timeframes. In general, 10 years seems to be an adequate timeframe for Greenland, since that corresponds to the time when pelagic fishing began in Eastern Greenland. The reason for these differing timeframes could be caused by approaching the ISA in different ways and using different interpretations.

The main body of this section focuses on Part B – getting the information (Gregory et al., 2023). The section starts with a representation of the process of selecting indicators, followed by the process of completing the Adjacency and Sensitivity matrix information, which was later used to inform the Kumu software. The last part will address Part C – Using the information and will present the Kumu results. This will be followed by a summary and general evaluation part.

Overview of all indicators

The process of selecting indicators entailed a broad consideration of the system in focus as well as its subsystems. The identification of indicators started with a practitioner ´brainstorming´ session of relevant indicators, followed by an elimination process, based on the importance and relevance of the indicator and other factors concerning data availability and data quality. For example, indicators with time-series that did not overlap were not considered in the final Kumu application. Indicators, such as marine diversity and pelagic fish abundance index, that did not overlap in space were also not further considered. Furthermore, the qualitative nature of certain indicators made it difficult to represent the indicator quantifiably in the ISA, despite their importance for pelagic fisheries – for instance, international agreements and transparency and inclusion in political decision-making. In addition, paying attention to the specific ecological and societal contexts affects the management options of the System in Focus. This means that elements of goods and benefits, activities, and needs are closer to area-specific, whereas ecosystem services, marine processes and functioning, and pressures are Arctic-site-specific, or more accurately, pelagic-ecosystem-specific. Table 13 gives a list of all indicators, with information on which ones were included by site within the Arctic DA in the final Kumu application.

Table 13. Overview of all indicators considered, and the ones finally included in Kumu by country within the Arctic Demonstration Area. GDP: Gross Domestic Product.

Type	Indicator		Included into Kumu			Source
			Iceland	Faroe Islands	Greenland	
Goods and Benefits	Income and Employment	Total employment fishing/processing and aquaculture	x			Statistics Iceland (2023)
		Total employment in fisheries		x		Statistics Faroe Islands (2023)
		Total employment in fisheries			x	Greenland Statistics (2023)
	Economic Contribution	Profitability				Indicator considered, but not
		Export value of pelagic products	x			Statistics Iceland (2023)
		GDP fishing (pelagic) and aquaculture	x			Statistics Iceland (2023)
		Export value herring and mackerel			x	Greenland Statistics (2023)
		Fisheries contribution to GDP			x	Greenland Statistics (2023)
		GDP pelagic		x		Statistics Faroe Islands (2023)
		GDP (aquaculture)		x		Statistics Faroe Islands (2023)

		GDP (demersal)		x		Statistics Faroe Islands (2023)
	Biodiversity	Low Arctic marine species diversity index				CAFF (2012)
Ecosystem services	Food Provision	Salmon from aquaculture	x	x		Statistics Iceland (2023), Statistics Faroe Islands (2023)
		Total catch		x		Statistics Faroe Islands (2023)
		Catch mackerel	x	x	x	ICES (2023a), Statistics Faroe Islands (2023)
		Catch blue whiting	x	x	x	ICES (2023a), Statistics Faroe Islands (2023)
		Catch herring	x	x	x	ICES (2023a), Statistics Faroe Islands (2023)
		Catch capelin	x			MFRI (2023)
		Catch demersal				Indicator considered, but not included in the end
	Ecosystem resilience / Biodiversity maintenance	Number of seabirds		x		Marine Research Faroese (2023)
		Pelagic fish abundance index				CAFF (2012)
		Net primary production	x	x	x	ICES (2023b), Oregon State University
		Zooplankton				Indicator considered, but not included in the end due to limited data sources
		Whale counts				NAMMCO (2019)
	Community Development	Population by municipality	x			Statistics Iceland (2023)
		Population East Greenland			x	Greenland Statistics (2023)
Marine Processes and Functionin g	Reproductive success	SSB mackerel	x	x	x	ICES (2023a)
		SSB blue whiting	x	x	x	ICES (2023a)
		SSB herring	x	x	x	ICES (2023a)
		SSB capelin	x		x	MFRI (2023)
		Marine mammal				MFRI (internal)
		Seabirds				Indicator considered, but not included in the end due to limited data sources
	Fish migration and behaviour	Mackerel occurrence per country	x	x	x	Anon. (2023a)
		Herring occurrence per country	x	x	x	Anon. (2023c)
		Blue whiting occurrence per country	x	x	x	Anon. (2023b)
Dominant pressures	Fishing pressure	Mackerel	x	x	x	ICES (2023a)
		Blue whiting	x	x	x	ICES (2023a)
		Herring	x	x	x	ICES (2023a)
		Capelin	x			ICES (2023a)
		Cod		x		ICES (2023c)
	Greenhouse gas emissions	Carbon dioxide emissions	x		x	SFS (2023)
		Fuel consumption in fisheries		x		Statistics Faroe Islands (2023)
	Waste	Input of substances				MFRI (internal)
	Sustainability marketing	Increased competition				Indicator considered, but not included in the end due to limited data sources
	Product innovations	Number of products				Indicator considered, but not included in the end due to limited data sources
	General disturbance	Noise and movement				Indicator considered, but not included in the end due to limited data sources
Dominant activities	Pelagic fishing	Days at sea	x			MFRI (internal)
		Number of vessels	x	x	x	MFRI (internal), Faroe Islands Fisheries Inspection (2023), Greenland Statistics
		Pelagic sale quantity		x		Statistics Faroe Islands (2023)

	Demersal fishing	Demersal sale quantity		x		Statistics Faroe Islands (2023)
	Shipping	Transport through seaports	x			Statistics Iceland (2023)
	Port and harbour operation	Arrived ships		x		Statistics Faroe Islands (2023)
	Cruise ships					Statistics Iceland (2023)
	Technology development					Statistics Iceland (2023)
	Aquaculture	Smolts (salmon) put in sea		x		Statistics Faroe Islands (2023)
Needs/ Drivers	Food security	Population growth	x	x	x	Statistics Iceland (2023), Statistics Faroe Islands (2023) Greenland Statistics (2023)
	Identity and belonging	National workers in fishing and aquaculture	x			Statistics Iceland (2023)
		National workers in fishing		x		Statistics Faroe Islands (2023)
		National workers in fish processing		x		Statistics Faroe Islands (2023)
	Self-actualization	Total employed in fisheries, processing and			x	Greenland Statistics (2023)
		Unemployment	x	x	x	Statistics Iceland (2023), Statistics Faroe Islands (2023)
		Job vacancies in fishing and aquaculture	x			Statistics Iceland (2023)
		Energy usage				Indicator considered, but not included in the end due to limited data sources
	Healthier climate	Fuel consumption total		x		Statistics Faroe Islands (2023)
	Esteem	Fish price				Statistics Iceland (2023)
		Business investment forecast				Statistics Iceland (2023)

Adjacency and Sensitivity Matrices

Once the final indicators were identified, data for each indicator was registered in the BOTs, followed by the process of filling in the Adjacency and Sensitivity Matrices. These were, however, not straightforward to complete. According to the guidance, the determination of assessing a +/- relationship (+ being in the same direction, - being in the opposite direction) was dependent upon the graphs in the BOTs. This led to several concerns.

First, a connection represented by + can easily be interpreted as a general increasing trend and – as a general decreasing trend, whereas they determine whether the relationships follow the same trend (either increasing or decreasing) or not (increasing and decreasing or vice versa). This impression is further strengthened due to the usage of the colour green as a representation of a + and red for -. To avoid false interpretations, the users urge the developers to adapt the colour scheme in the matrices as well as the Kumu application.

Second, it was often difficult to identify any distinct trend in BOTs, since some indicators undergo many fluctuations. Hence, the users added linear trendlines to aid in identifying a trend. This can, however, also completely remove any trend and let the BOT trend appear to be unchanged. Additionally, it is difficult to compare graphs that have different units and scales. In the BOTs of the adjacency matrices sheets, some graphs do not appear to have any trend, whereas there may be a more distinct trend with the original scale of the indicator. Furthermore, drawing a relationship between two trends without a time series analysis can also be misleading. For example, two trends, each with a U-shape over time, would each be considered to have a flat trend over time and no relationship, but in time series analysis would be highly correlated. Usually, removing constant trends over time is the first step of a time series analysis to avoid spurious correlations among variables.

Third, forcing a relationship between certain indicators that do not have an evidence-based connection/relationship, can result in misleading results or false interpretations by stakeholders later. Although some would argue that, theoretically, elements are connected one way or another, to assess this in actual practicalities with only a few indicators and only in a +/- relation, could result in oversimplifying the phenomenon. Another reason for this concern could be that the selected indicators were not appropriate for this exercise and would require another elimination process of indicators with a more in-depth focus on pelagic fisheries.

Fourth, following the BOTs in assessing the +/- relationship sometimes resulted in conflicting outcomes. For instance, considering the total catches of three species (mackerel, herring, and blue whiting) related to their Spawning Stock Biomass (SSB), should, theoretically, follow the same pattern (i.e., the catch should increase with SSB). Despite this, according to the matrices, e.g., in the case of mackerel and herring, this functional relationship is not observed. Consequently, the observed trends are affected by the selection of the timeframe, indicators, or external factors, and therefore the BOTs do not always capture complex functional relations between indicators. In this regard, it is important to distinguish between the goals of this exercise. If the point of creating 'causal loop diagrams' is to infer causal relationships, then the process of drawing causal relationships from temporal trends in the BOTs can be very misleading. If the point is instead to create a depiction of 'current trends and status', then the process would make more sense, but then the name 'causal loop diagram' is misleading. If the goal is to create causal relationships in a 'causal loop diagram', then perhaps it would have been better to make causal diagrams from well-supported theoretical relationships rather than beginning with indicators and BOTs.

Following up on this, assessing the strength of connections and relationships, was more guesswork than built upon certainty. On the one hand, the guidance does not provide any criteria on how to qualify connections/relationships as weak, medium, or strong +/- relations if following the BOTs, increasing the risk of subjective qualifications of the connections. On the other hand, the BOTs do not always necessarily prove efficient in informing a connection/relationship, and as such, it would be very useful with the option of allocating a 0 or an unclear relationship between indicators as some indicators may be very relevant to the system but without a connection to other indicators in this exercise. This would also require the matrices to be completed with support from local knowledge and expert judgement.

To illustrate further the above concerns, Tables 14-31 represent the matrices for Iceland and the Faroe Islands. These tables include information on the +/- relationship between indicators following the BOTs, the strength of these relationships according to the exercises, as well as additional information on whether these connections make sense according to the expert knowledge of the DA or not. The last step is done by giving relations a number from 1—3, that have the following meaning:

1 = No relationship between indicators in reality
2 = Mismatch between BOTs data and expert assessment
3 = Uncertain connection with possibly correct BOTs

The numbers are allocated according to the local knowledge and expert judgment on the indicators and local contexts. This included the involvement of the Arctic DA experts, who assessed the allocation of the numbers. The experts came from various academic backgrounds, including fisheries scientists, ecological biologists, and social scientists, which allowed for a comprehensive judgment on these matters.

Goods and Benefits vs. Ecosystem Services

The links between ecosystem services and benefits and goods, in the three areas within the Arctic can be seen in Tables 14-16. An example of a mismatch or misleading interpretation of the BOT results/comparisons is the relationship between total employment and total catch by species. In the Faroese case, the BOTs indicate that total employment decreases when catch increases. Theoretically, the more catches, the more work, the more employment. The trends are therefore expected to follow the same direction, as in the case of Iceland regarding mackerel and herring. However, one can also question whether there is a connection/relationship between these two indicators at all, as a decrease in employment and an increase in catches could be explained by other factors, such as the efficiency of the fleet, processing, and/or degree of exportation of raw materials. Therefore, the indicators can also represent an uncertain connection.

Table 14. Links between ecosystem services and goods and benefits, in Iceland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, GDP = Gross domestic product, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Goods and Benefits		
		Total employment fishing/processing and aquaculture	Export value of pelagic products	GDP fishing and aquaculture
Ecosystem Services	Total catch MAC	Weak + [3]	Weak +	Weak - [2]
	Population by municipality	Medium - [2]	Medium - [1]	Medium + [3]
	Net primary production (NorSea)	Weak - [1]	Weak - [1]	Weak + [1]
	Salmon from aquaculture	Medium – [2; 3]	Weak - [1]	Medium + [3]
	Total catch HER	Weak + [3]	Weak +	Weak - [2]
	Total catch BW	Medium – [2; 3]	Medium - [2]	Medium +
	Total catch CA	Medium + [3]	Medium +	Medium - [2]

Table 15. Links between ecosystem services and goods and benefits, in Faroe Islands. Note: MAC = mackerel, BW = blue whiting, HER = herring, GVA = Gross Value Added, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Goods and benefits				
		Total employment (fisheries)	GVA (Fisheries)	Pelagic exports	Aquaculture exports	Demersal exports
Ecosystem services	Salmon production (aquaculture)	Weak - [1]	Weak + [1]	Weak +	Strong +	Weak - [1]
	Seabirds	Weak + [1]	Weak - [1]	Weak - [1]	Weak - [1]	Weak + [1]
	Primary production (NorSea)	Weak + [1]	Weak - [1]	Weak - [1]	Weak - [1]	Weak + [1]
	Total catch MAC	Weak - [2; 3]	Strong +	Strong +	Weak + [1]	Weak - [1]
	Total catch BW	Weak - [2; 3]	Med +	Med +	Weak + [1]	Weak - [1]
	Total catch HER	Weak – [2; 3]	Strong +	Strong +	Weak + [1]	Weak - [1]

Table 16. Links between ecosystem services and goods and benefits, in Greenland. Note: MAC = mackerel, BW = blue whiting, HER = herring, GDP = Gross Domestic Product, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Goods and Benefits		
		Total employment fishing, processing and trade	Export value of pelagic products	GDP contribution fishing
Ecosystem Services	Total catch MAC	Weak + [3]	Weak -	Weak - [2]
	Population by municipality	Medium - [3]	Medium + [1]	Medium - [3]
	Net primary production (NorSea)	Weak - [1]	Weak + [1]	Weak + [1]
	Total catch HER	Weak + [3]	Weak -	Weak - [2]
	Total catch BW	Medium – [2; 3]	Weak + [2]	Weak +

Ecosystem Services vs. Marine processes and functioning

As illustrated in the matrices for all three countries for Ecosystem Services and Marine processes and functioning (Tables 17-19), in reality, most cases are considered to have no relation/connection at all but may still be important for the ecosystem as a whole. Considering indicators such as primary production, these do not necessarily have any direct connection with pelagic species broadly speaking. This is shown by work done by the ICES WGNOR working group which has tested for relationships while developing data-intensive state-of-the-art integrated ecosystem assessments. However, primary production does serve as the fundamental element of the food web and is therefore rather important. Recently, there has been an increase in the literature trying to understand the connection between primary production, zooplankton, and pelagic fish stocks – and the users know there must be some kind of connection – but so far studies have failed to understand the exact relationships composing these complex connections. Additionally, life stages more closely connected to primary production, such as recruitment, occur outside of the Arctic DA ecosystem (further to the south).

Additionally, there should not be any connection between the different pelagic species, indicating that no causal relationship should be detailed. There have been studies of the relationship between mackerel and herring as both species eat the same zooplankton species, but it has been shown that they are opportunistic feeders and just feed on what is available. Moreover, both species have a slightly lagged migration time and do not overlap much.

Lastly, and importantly, the various BOTs results concerning pelagic species are a fundamental example of the BOTs inability to capture the complexity of reality. For instance, whilst total catch should be positively connected with SSB and occurrence level for these species, according to the BOTs, this is only the case for blue whiting. Hence, the mismatch in the Faroe Islands case is concerning for the other two species for which occurrence data were available (mackerel and herring). Similar mismatches are observed in the Greenlandic BOTs for the same indicators.

Table 17. Links between ecosystem services and marine processes and functioning, in Iceland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Ecosystem Services						
		Total catch MAC	Population by municipality	Net primary production (NorSea)	Salmon from aquaculture	Total catch HER	Total catch BW	Total catch CA
Marine processes and functioning	SSB MAC	Weak +	Medium - [1]	Weak - [1]	Weak - [1]	Weak + [1]	Medium - [1]	Weak - [1]
	MAC occurrence (EEZ)	Medium +	Medium - [2]	Weak - [1]	Medium - [1]	Medium + [1]	Medium - [1]	Medium - [1]
	SSB BW	Weak + [1]	Medium + [1]	Weak + [1]	Weak + [1]	Weak - [1]	Medium +	Weak + [1]
	SSB HER	Weak - [1]	Medium - [1]	Weak - [1]	Weak - [1]	Weak +	Weak - [1]	Weak + [1]
	SSB CA	Weak + [1]	Weak + [1]	Weak + [1]	Weak + [1]	Weak - [1]	Medium + [1]	Medium - [1]
	HER occurrence (EEZ)	Weak + [1]	Weak +	Weak + [1]	Weak + [1]	Weak - [2]	Weak + [1]	Weak - [1]
	BW occurrence (EEZ)	Weak + [1]	Weak +	Weak + [1]	Weak + [1]	Weak + [1]	Weak +	Weak - [1]

Table 18. Links between ecosystem services and marine processes and functioning, in Faroe Islands. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Ecosystem services					
		Salmon production aquaculture	Number of seabirds	Primary production (NorSea)	Total catch MAC	Total catch BW	Total catch HER
Marine processes and functioning	MAC occurrence (EEZ)	Weak - [1]	Weak + [1]	Weak + [1]	Med - [2]	Weak -[1]	Weak -[1]
	SSB MAC	Weak - [1]	Weak + [1]	Weak + [1]	Med - [2]	Weak -[1]	Weak - [1]
	SSB BW	Weak + [1]	Weak + [1]	Weak + [1]	Weak + [1]	Med +	Weak +[1]
	SSB HER	Weak - [1]	Weak + [1]	Weak + [1]	Weak - [1]	Weak - [1]	Med - [2]
	BW occurrence (EEZ)	Weak + [1]	Weak - [1]	Weak – [1]	Weak + [1]	Med +	Weak +
	HER occurrence (EEZ)	Weak - [1]	Weak + [1]	Weak + [1]	Weak - [1]	Weak -[1]	Med – [2]

Table 19. Links between ecosystem services and marine processes and functioning, in Greenland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Ecosystem Services				
		Total catch MAC	Population by municipality	Net primary production (NorSea)	Total catch HER	Total catch BW
Marine processes and functioning	SSB MAC	Weak +	Weak[+1]	Weak - [1]	Weak + [1]	Weak - [1]
	MAC occurrence (EEZ)	Medium +	Weak+ 1	Weak - [1]	Medium + [1]	Weak - [1]
	SSB BW	Weak - [1]	Weak - [1]	Weak + [1]	Weak - [1]	Weak +
	SSB HER	Weak - [1]	Weak + [1]	Weak - [1]	Weak +	Weak - [1]
	HER occurrence (EEZ)	Weak + [1]	Weak +	Weak + [1]	Weak - [2]	Weak - [1]
	BW occurrence (EEZ)	Weak + [1]	Weak +	Weak + [1]	Weak + [1]	Weak - [2]

Marine processes and functioning vs. Pressures

The case of the Marine processes and functioning vs. Pressures matrix (Tables 20-22) illustrates the misleading interpretation of the BOTs very well. First (and similar to other matrices), there should be no connection/relationship at all between the fishing pressures of the different species as they are caught in one fishery but at different times of the year. There is also not a lot of by-catch in the pelagic fishery and no bottom contact, hence it is not considered as damaging to the ecosystem as the demersal fishery, which is also a mixed fishery. Second, although fishing mortality is an important factor to consider in this system, it is a man-made variable and therefore should theoretically not be connected to the SSB of any of the species. In this case, either no connection should be put into the matrices, or the connection should be based on theory only, not on the BOT comparison.

Table 20. Links between pressures and marine processes and functioning, in Iceland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Marine processes and functioning						
		SSB MAC	MAC occurrence (EEZ)	SSB BW	SSB HER	SSB CA	HER occurrence (EEZ)	BW occurrence (EEZ)
Pressures	CO_2 emissions	Weak + [3]	Weak + [3]	Weak - [3]	Weak + [3]	Weak - [3]	Weak - [3]	Weak - [3]
	Fishing mortality MAC	Weak - [1]	Weak - [1]	Weak + [1]	Weak - [1]	Medium + [1]	Weak + [1]	Weak + [1]
	Fishing mortality BW	Medium - [1]	Medium - [1]	Weak + [1]	Weak - [1]	Medium + [1]	Weak + [1]	Weak + [1]
	Fishing mortality HER	Medium - [1]	Medium - [1]	Weak + [1]	Medium - [1]	Medium + [1]	Medium + [1]	Weak + [1]
	Fishing pressure CA	Medium + [1]	Medium + [1]	Medium - [1]	Medium + [1]	Medium - [1]	Medium - [1]	Weak - [1]

Table 21. Links between pressures and marine processes and functioning, in Faroe Islands. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Marine process and Functioning					
		Mackerel occurrence (EEZ)	SSB MAC	SSB BW	SSB HER	BW occurence (EEZ)	HER occurence (EEZ)
Pressures	Fuel consumption (fisheries)	Med + [3]	Med + [3]	Med + [3]	Weak + [3]	Weak + [3]	Med + [3]
	Fishing mortality MAC	Med – [1]	Med - [1]	Weak - [1]	Weak - [1]	Weak - [1]	Weak - [1]
	Fishing mortality BW	Weak + [1]	Weak + [1]	Med + [1]	Weak - [1]	Weak + [1]	Weak - [1]
	Fishing mortality HER	Weak - [1]	Weak - [1]	Weak - [1]	Weak + [1]	Weak - [1]	Med - [1]
	Fishing mortality COD	Weak - [1]	Weak - [1]	Weak – [1]	Weak - [1]	Weak - [1]	Weak - [1]

Table 22. Links between pressures and marine processes and functioning, in Greenland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, EEZ: Exclusive Economic Zone, SSB: Spawning Stock Biomass, NorSea = Norwegian Sea, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Marine processes and functioning					
		SSB MAC	MAC occurrence (EEZ)	SSB BW	SSB HER	HER occurrence (EEZ)	BW occurrence (EEZ)
Pressures	CO_2 emissions	Weak - [3]	Medium - [3]	Weak - [3]	Weak - [3]	Weak + [3]	Weak - [3]
	Fishing mortality MAC	Weak - [1]	Medium - [1]	Weak + [1]	Weak - [1]	Weak + [1]	Weak + [1]
	Fishing mortality BW	Medium - [1]	Medium - [1]	Weak + [1]	Weak - [1]	Weak + [1]	Weak - [1]
	Fishing mortality HER	Weak - [1]	Medium - [1]	Weak + [1]	Medium - [1]	Weak + [1]	Weak - [1]

Pressures vs. Activities

The links between activities and pressures, in the three areas within the Arctic can be seen in Tables 23-25. In addition to similar issues already mentioned, especially the negative relationship between the number of companies, transport through seaports, and carbon dioxide emissions of fishing vessels in the Icelandic Pressure vs. Activities matrix are apparent as there should be no connection at all between those indicators, or if a relationship does exist, then a positive relationship would be expected.

Table 23. Links between pressures and activities, in Iceland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Pressures				
		CO_2 emissions	Fishing mortality MAC	Fishing mortality BW	Fishing mortality HER	Fishing pressure CA
Activities	Total effort	Weak +	Weak - [1]	Medium - [1]	Medium - [1]	Medium + [1]
	Number of companies	Weak - [2]	Weak + [1]	Medium + [1]	Medium + [1]	Medium - [1]
	Number of vessels	Weak +	Weak - [1]	Medium - [1]	Medium - [1]	Weak + [1]
	Transport through seaports	Weak - [2]	Weak + [1]	Medium + [1]	Weak + [1]	Medium - [1]

Table 24. Links between pressures and activities, in Faroe Islands. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Pressures				
		Fuel consumption (fisheries)	Fishing Mortality MAC	Fishing Mortality BW	Fishing Mortality HER	Fishing Mortality COD
Activities	Pelagic sale quantity	Med - [3]	Med + [1]	Med + [1]	Med + [1]	Med - [1]
	Demersal sale quantity	Weak + [3]	Weak -[1]	Weak -[1]	Weak +[1]	Weak -[1]
	Pelagic vessels	Weak - [2]	Weak +[1]	Weak - [1]	Weak +[1]	Weak -[1]
	Arrived ships	Med +	Weak – [1]	Weak + [1]	Weak - [1]	Weak -[1]
	Smolts put in sea	Weak - [3]	Weak + [1]	Weak – [1]	Weak + [1]	Weak +[1]

Table 25. Links between pressures and activities, in Greenland. Note: MAC = mackerel, BW = blue whiting, HER = herring, CA = Capelin, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Pressures			
		CO_2 emissions	Fishing mortality MAC	Fishing mortality BW	Fishing mortality HER
Activities	Number of companies	Weak +	Weak - [1]	Medium + [1]	Medium + [1]
	GL ships, catch	Weak -	Weak -	Medium - [1]	Medium -

Activities vs. Drivers/Needs

Mismatches in the activities and driver matrices (Tables 26-28) are, for example, the proportion of Icelandic nationals working in fisheries, job vacancies in fishing and aquaculture, and the number of companies, which is negatively related according to the BOTs but would normally be expected to be positive. Similarly, with increased unemployment, the days at sea and number of vessels are expected to decrease, not increase. There should be no connection between nationals working in the fishing and aquaculture sector and arrived ships or transport through seaports. These are important activities in the pelagic environment, but the ships arriving in Icelandic or Faroese harbours are not necessarily influenced by the proportion of nationals working in the fishing and aquaculture sectors. In the Greenlandic case, the mismatches are mostly concerning the relationship between the indicators number of people employed in fisheries and national unemployment numbers versus the indicator number of companies. One would expect them to have a positive connection.

Table 26. Links between drivers and activities, in Iceland. Note: [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Activities			
		Days at sea	Number of companies	Number of vessels	Transport through seaports
Drivers	% nationals working in fisheries	Medium + [3]	Medium – [2; 3]	Medium + [3]	Weak - [3]
	Job vacancies fishing and aquaculture	Weak + [3]	Weak – [2; 3]	Weak + [3]	Weak - [3]
	Unemployment	Weak + [3]	Weak - [3]	Weak + [2; 3]	Weak - [3]
	Population	Weak - [3]	Medium + [3]	Medium – [2; 3]	Weak + [3]

Table 27. Links between drivers and activities, in Faroe Islands. Note: [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Activities				
		Pelagic sale quantity	Demersal sale quantity	Pelagic vessels	Arrived ships	Smolts put in sea
Drivers	Population growth rate	Weak + [3]	Weak - [3]	Weak + [1]	Weak - [1]	Weak + [3]
	Nationals working in fishing ships	Med – [3]	Med + [3]	Weak - [3]	Weak – [1]	Med - [1]
	Nationals working in fish processing	Med - [1]	Med + [3]	Weak – [1]	Weak – [1]	Med - [1]
	Fuel consumption total	Med + [3]	Med - [3]	Med +	Weak – [2; 3]	Med +
	Unemployment rate	Med - [3]	Med + [3]	Weak – [1]	Weak + [1]	Med -

Table 28. Links between drivers and activities, in Greenland. Note: [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Activities	
		Number of companies	Greenlandic ships, catch
Drivers	Employed in fishing	Medium – [2]	Medium + [3]
	Unemployment	Weak – [2,3]	Weak + [3]
	Population	Weak - [3]	Weak + [3]

Drivers/Needs vs. Goods and Benefits

An obvious mismatch between the Drivers and Goods and benefits BOTs occurs when comparing the unemployment rate or export value of pelagic products and the total employment rate in fishing/processing and aquaculture (Tables 29-31). This mismatch is also observed in the Greenlandic case. With an increase in unemployment, a decrease in employment rate (i.e. a negative relationship) would be expected, not a positive relationship. Similarly, with an increase in GDP in fishing and aquaculture a positive influence would also be expected on job vacancies, not negative.

Table 29. Links between drivers and goods and benefits, in Iceland. Note: GDP: Gross Domestic Product, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Iceland		Drivers			
		% nationals working in fisheries	Job vacancies fishing and aquaculture	Unemployment	Population
Goods and Benefits	Total employment fishing/processing and aquaculture	Weak + [3]	Weak +	Weak + [2]	Weak - [3]
	Export value of pelagic products	Weak + [3]	Weak + [3]	Weak + [2]	Weak - [2]
	GDP fishing and aquaculture	Weak - [2]	Weak - [2]	Weak - [3]	Medium + [3]

Table 30. Links between drivers and goods and benefits, in Faroe Islands. Note: GVA: Gross Value Added, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Faroe Islands		Drivers				
		Population growth rate	Nationals working in fishing vessels	Nationals in fish processing	Fuel consumption total	Unemployment rate
Goods and benefits	Tot employment (Fisheries)	Med – [3]	Weak + [3]	Weak + [3]	Weak - [1]	Med + [2; 3]
	GVA (Fisheries)	Med +	Med - [3]	Med - [3]	Weak + [3]	Med - [3]
	Pelagic export	Med +	Med - [3]	Med - [3]	Weak + [3]	Med - [3]
	Aquaculture export	Med +	Weak - [1]	Weak - [1]	Weak + [3]	Med - [3]
	Demersal export	Med - [1]	Med - [3]	Med + [3]	Med - [2]	Weak + [3]

Table 31. Links between drivers and goods and benefits, in Greenland. Note: GDP: Gross Domestic Product, [1] = No theoretic relationship between indicators, [2] = Mismatch between BOTs data and expert assessment, [3] = Uncertain connection.

Greenland		Drivers		
		Employed in fishing	Unemployment	Population
Goods and Benefits	Total employment fishing, processing and trade	Medium +	Weak + [2]	Weak + [3]
	Export value of pelagic products	Medium - [3]	Medium - [3]	Weak - [3]
	Growth contribution to GDP by fishing	Weak - [2]	Weak - [3]	Weak - [3]

Kumu

Once the matrices were finalized, the relevant Excel sheets were exported and imported into the Kumu software. This import process was successful, but challenges concerning the computational demands of the Kumu software resulted in the system crashing when trying to detect loops (see legend in Figure 26). After various attempts and meetings with WP3 task leads, it was decided to remove relationships between indicators that were assessed to have no support of connection (e.g., the 1 = no theoretic relationship between indicators). This significantly reduced the number of connections/relationships and allowed for the automatic detection of loops on the Kumu software. This process will be explained in this part, starting with Iceland, followed by the Faroe Islands and Greenland.

Figure 26. Legend for the Kumu application.

Kumu results – Iceland

Figures 27 and 28, illustrate the overall results obtained using Kumu, both before and after removing connections with no theoretical support. In conducting the first loop analysis, the Icelandic results detected a total number of 496 loops (Figure 27). However, after adjusting and managing these relationships further, the number of loops reduced to 322 (Figure 28). In both cases, the classification of feedback loops as reinforcing (e.g., positive feedback) or balancing (e.g., negative feedback) becomes a very time-consuming process.

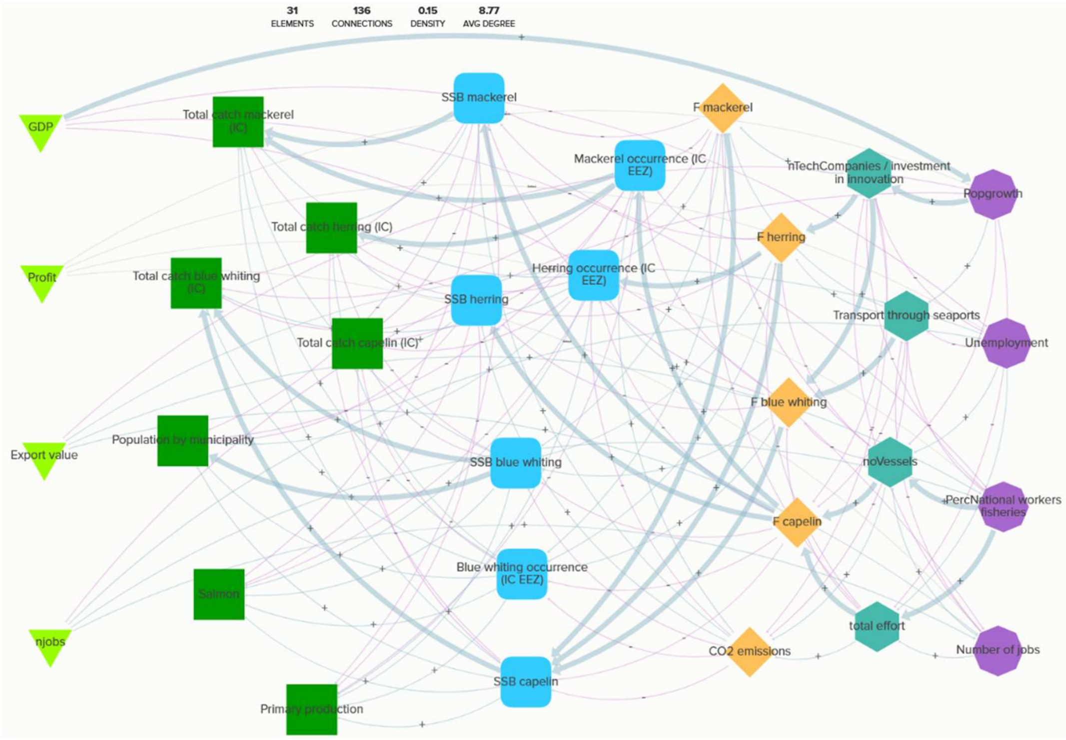

Figure 27. Kumu result for Iceland with all connections included as originally designed.

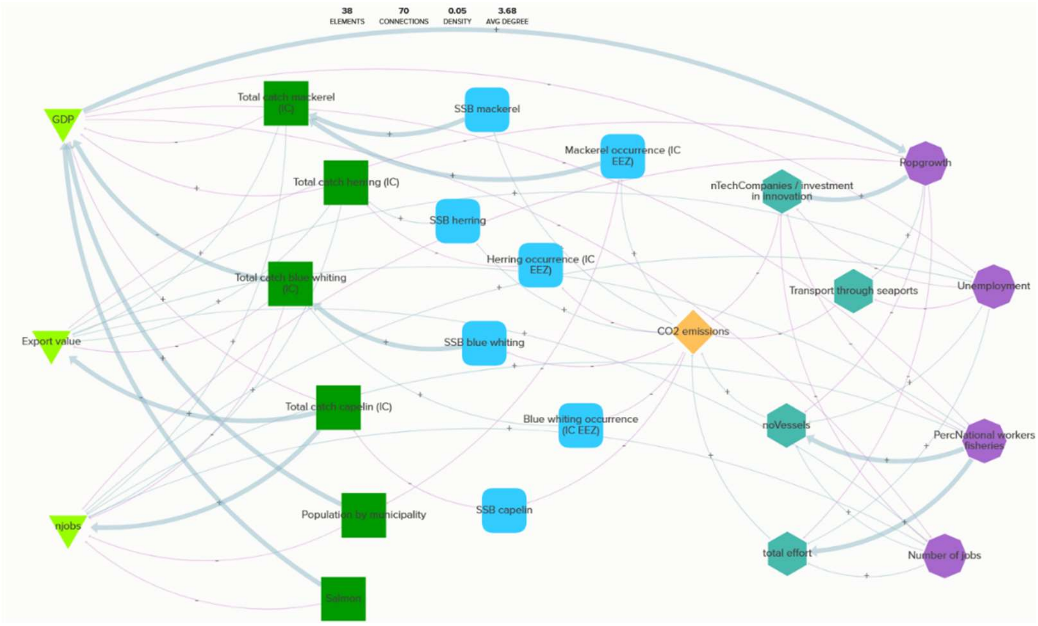

Figure 28. Kumu result for Iceland without connections marked as "1 = no theoretic relationship between indicators" in the matrices shown before. Kumu completely excludes indicators without connections, i.e. that may be important for the whole system, but not necessarily have a connection to other indicators (e.g. fishing pressures and primary production).

Figures 29-31 illustrate the most plausible examples of the loop analysis conducted in Kumu for the Icelandic region. In the first example (Figure 29), which is a positive feedback loop, an increase in Icelandic workers in the fishing sector would have a medium-strong impact on the pelagic fishing vessel total days at sea (total effort), increasing those. This would further lead to an increase in fishing vessel CO_2 emissions, which then, according to Kumu results, leads to an increase in Icelandic mackerel catch, further leading to an increase in the number of jobs in the fishing/processing and aquaculture sector, finally leading to an increase in Icelandic workers in the fishing sector. These relationships appear plausible, except that, in reality, it is unknown how the numbers of Icelandic workers in the fishing sector relate to total days at sea as well as how fishing vessel CO_2 emissions and mackerel occurrence would relate (whether the relationship is positive or negative).

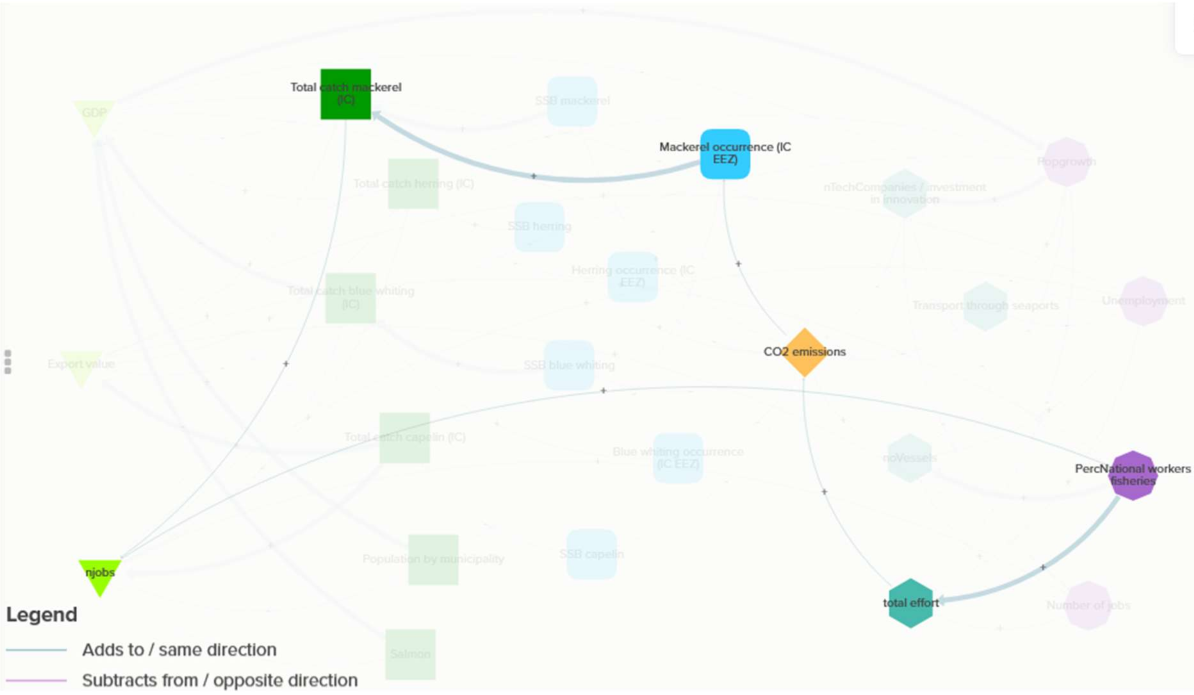

Figure 29. Kumu results for Iceland - Example 1 (positive feedback loop).

The second example (Figure 30) shows the following process: with an increase in the Icelandic population, days at sea (total effort) by the pelagic fishing vessels would decrease according to the causal loop. This would decrease CO_2 emissions by the fishing vessels, leading to a decrease of mackerel occurrence in the Icelandic EEZ, leading to an increase in the Icelandic population in the municipalities, increasing GDP for fishing and aquaculture, and finally having a positive effect on the Icelandic population. The relationship between Icelandic population growth and total effort is, however, unknown, as is the relationship between fishing vessels CO_2 emissions and mackerel occurrence. Further, a decrease in mackerel occurrence would not necessarily lead to an increase in the Icelandic population by municipality. It is well known that the contrary is the case, an increase in fish occurrence and catch helps maintain fish processing facilities, therefore offering more jobs and counteracting rural migrations (e.g. Kokorsch and Benediktsson, 2018).

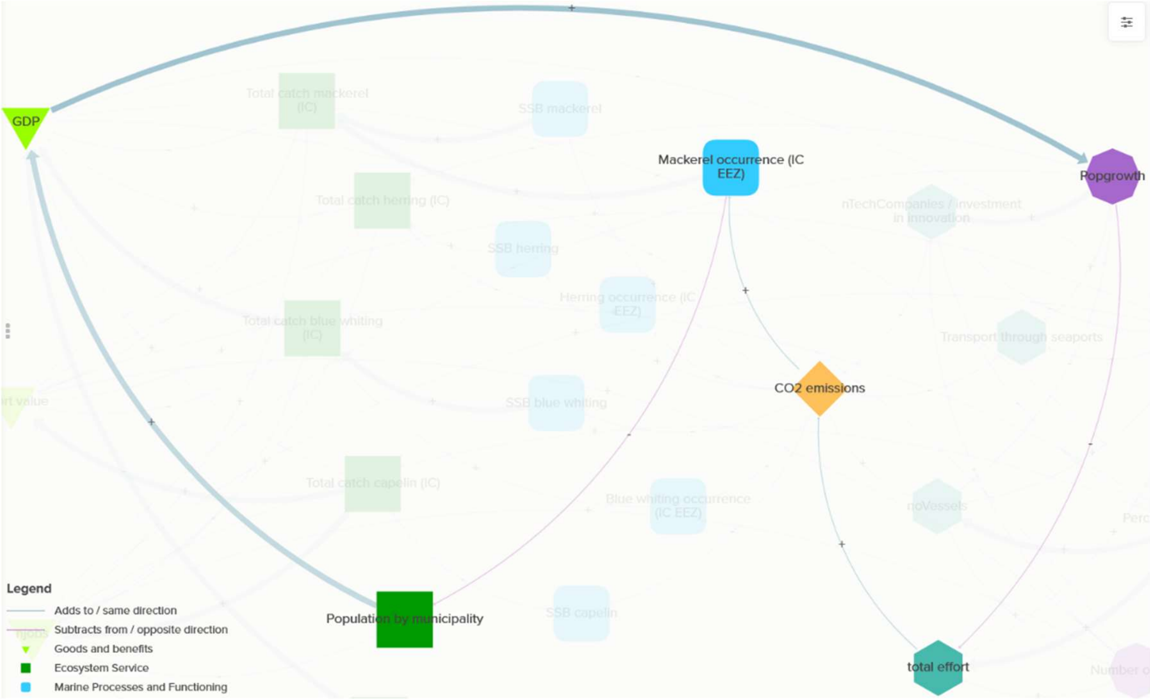

Figure 30. Kumu results for Iceland – Example 2.

The last example (Figure 31) is an example regarding the Icelandic herring fishery. An increase in GDP of the fishing and aquaculture sector leads to a decrease in overall employment, further leading to a decrease in total effort by the pelagic fleet, leading to a decrease in CO_2 emissions, a decrease in SSB, total Icelandic herring catch and finally also a decrease in GDP of the fishing and aquaculture sector. First, the relationship between unemployment and total effort is uncertain as is the relationship between CO_2 emissions and herring SSB (and if any relationship would exist a decrease in CO_2 emissions would be expected to increase, not a decrease of herring SSB). The causal loop results for the other pelagic species and socio-economic indicators are very similar to the examples shown here.

The last two examples illustrate that using observed relationships as interpreted from temporal trends to infer causal relationships will lead to unrealistic or erroneous loops within the causal loop diagram. Most likely, removing the relationships categorized as 2 ("Mismatch between BOTs data and expert assessment") will lead to more realistic, but far fewer, loops, as this will also remove a substantial number of relationships. Relationships categorized as 3 ("Uncertain connection with possibly correct BOTs") are more likely to show realistic loops within the causal loop diagram but should be assessed individually for their plausibility. Additional relationships could possibly be included, but without the support of observed indicator trends and instead supported by theoretical relationships or those demonstrated in published literature.

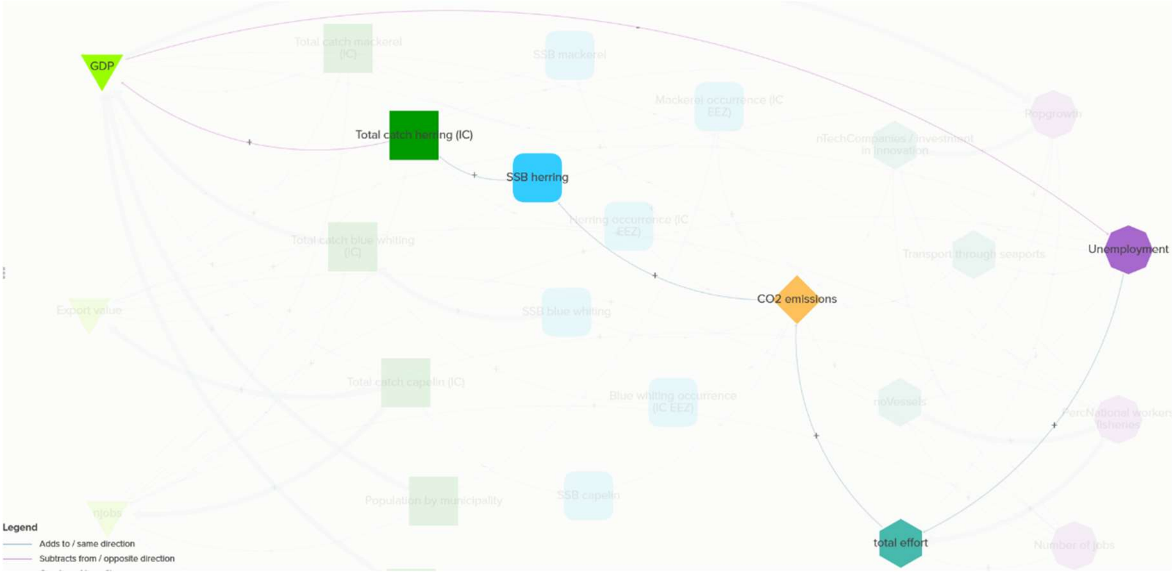

Figure 31. Kumu result for Iceland – Example 3.

Kumu results – The Faroe Islands

As with Iceland, the Faroe Islands also went through the process of reducing indicators from Kumu application once uploaded to the application. This is illustrated in Figures 32 and 33.

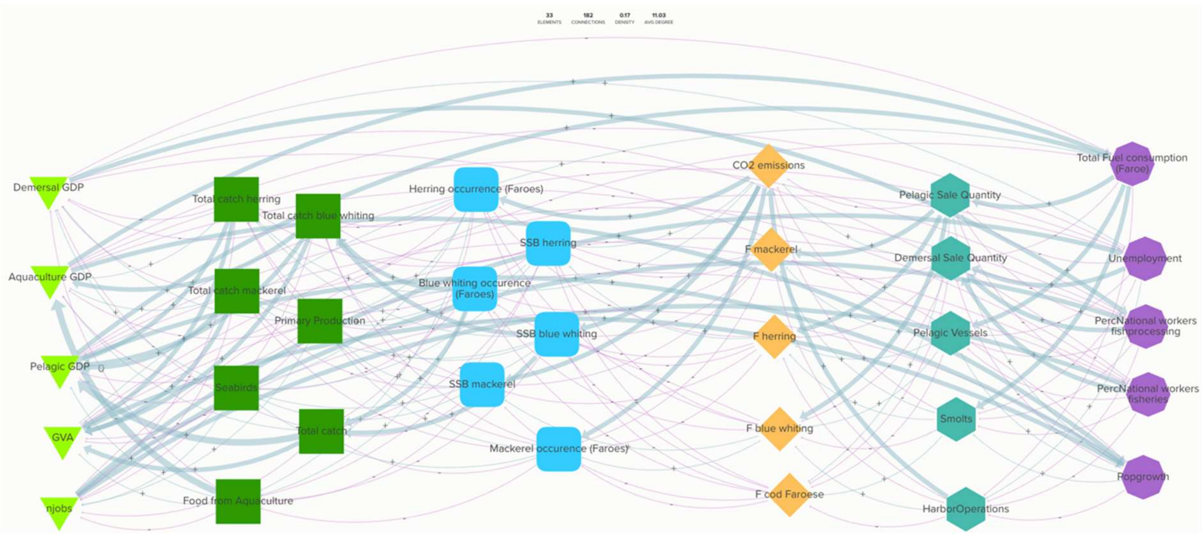

Figure 32. Kumu results for the Faroe Islands.

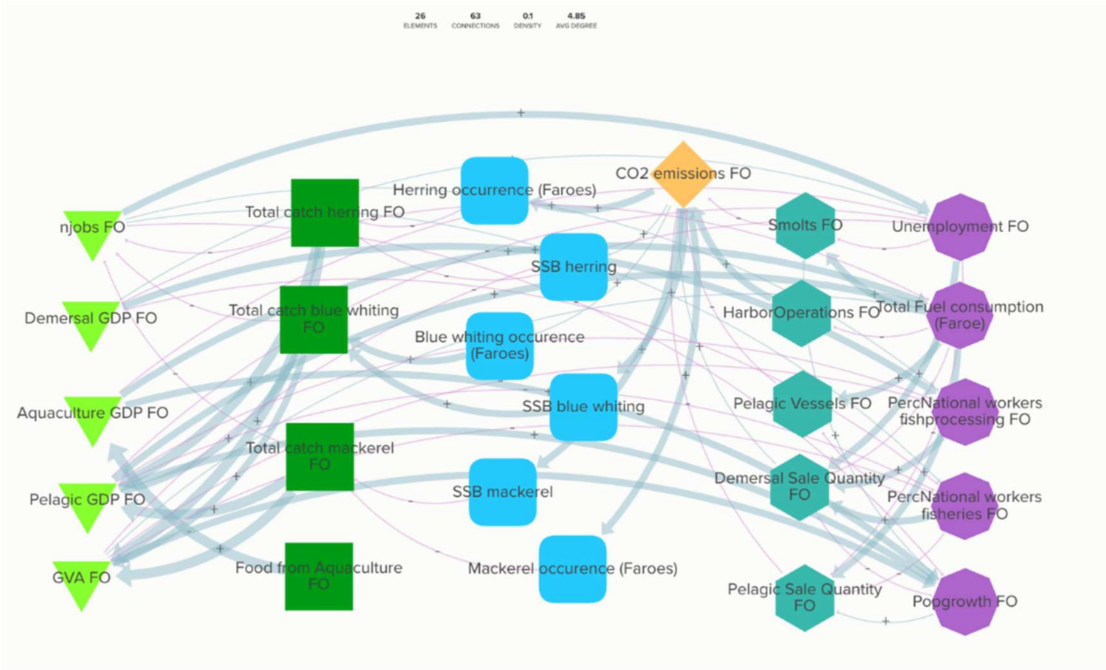

Figure 33. Kumu results for the Faroe Islands without connections marked as "1 = no theoretic relationship between indicators" in the matrices shown before.

Conducting the loop analysis (Figure 33) created a total of 240 loops in the case of the Faroe Islands. It can be added that in the first round of removing relationships/connections with no support, there were approximately 20,000 loops detected in the Kumu software in the case of the Faroe Islands. However, once adjusting and managing these relationships further, the number of loops reduced significantly. In both cases, the classification of feedback loops as reinforcing (e.g., positive feedback) or balancing (e.g., negative feedback) was a time-consuming process. This process (Figure 34) illustrates examples of a feedback loop and provides brief context and understanding of these loops.

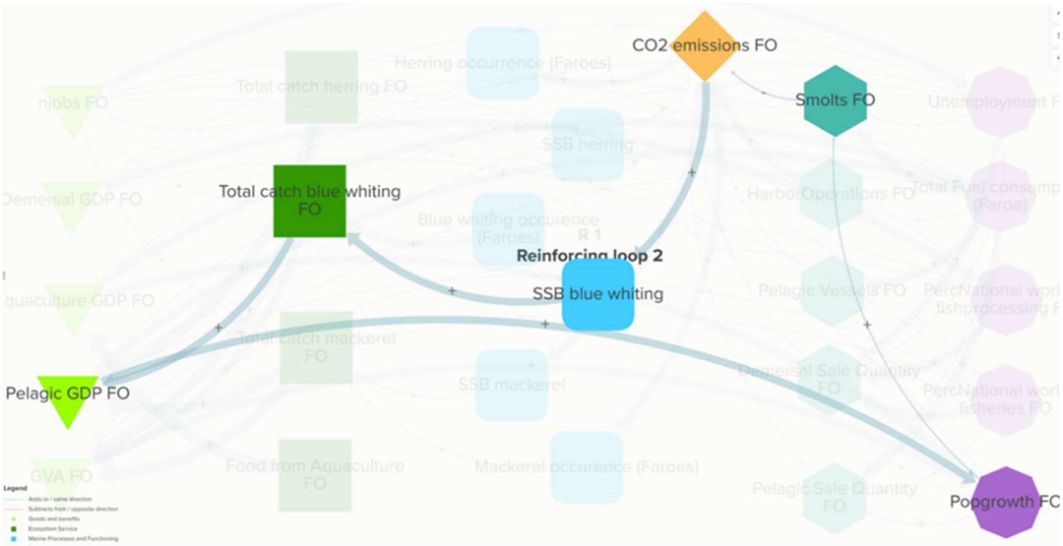

Figure 34. An example of a reinforcing loop in the Faroe Islands.

Figure 34 is an example of a positive/reinforcing loop and illustrates cause and effect in the following way: the economic contribution of pelagic fisheries (expressed in GDP) leads to population growth, which leads to more smolts being placed in the sea for aquacultural purposes, which decreases CO_2 emissions, increases SSB of blue whiting and catches of blue whiting, which leads to an increase in pelagic economic contribution. This feedback loop does, however, raise concerns about the logic of cause and effect. Although SSB levels, catch, economic contribution, and population growth positively interact in theory, the number of smolts placed in the sea for aquacultural activity and CO_2 emissions from fisheries, do not necessarily follow the logic of the loop, which is also expressed in the thickness of the arrows. Whether these should be understood as uncertain cause-and-effect elements, or as indicators that are too broad for the topic in focus, should be clarified. Considering this, it was also decided to try and establish a Kumu version that removed some indicators that do not directly have a give-to or take-from connection in the case of pelagic fisheries (e.g., demersal fisheries, aquaculture-related activities). This resulted in a further reduction of indicators, with Kumu's results represented in Figure 35.

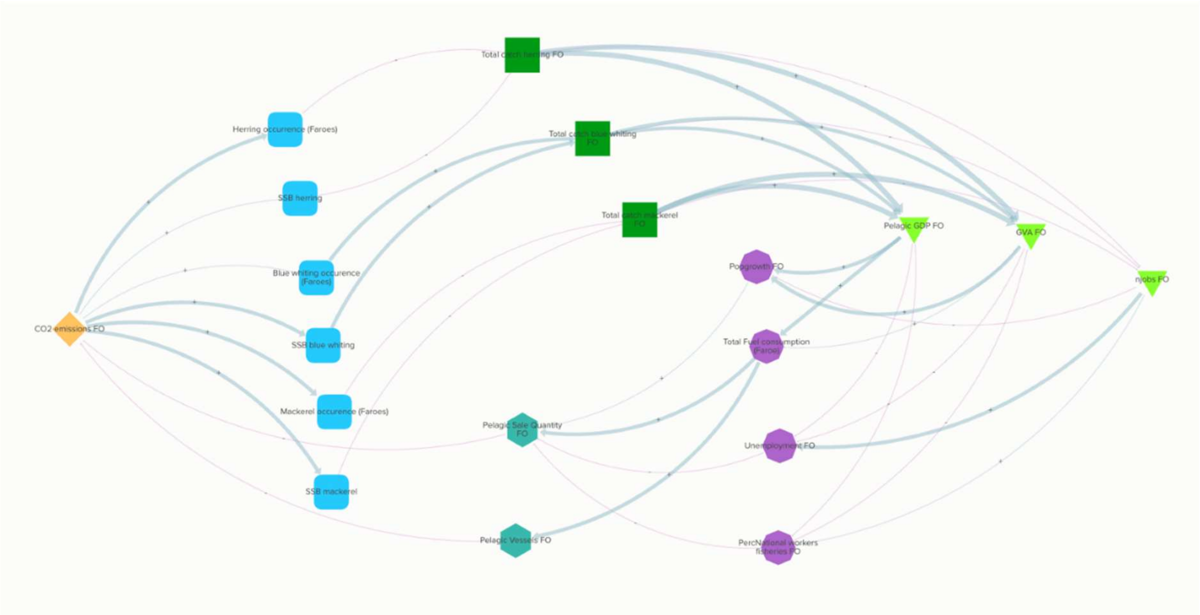

Figure 35. Kumu result for the Faroe Islands without connections marked as "1 = no theoretic relationship between indicators" and removal of other industries (e.g., demersal and aquaculture).

As Figure 35 illustrates, the logic of the Kumu results is becoming clearer. If looking at indicators for marine processes and functioning (e.g., occurrence and SSB by species), it is apparent that these have the same arrows coming to and going out. This arguably means that the indicators are ´´two sides of the same coin´´. Therefore, we can remove one of them, and the Kumu results will look like Figure 36.

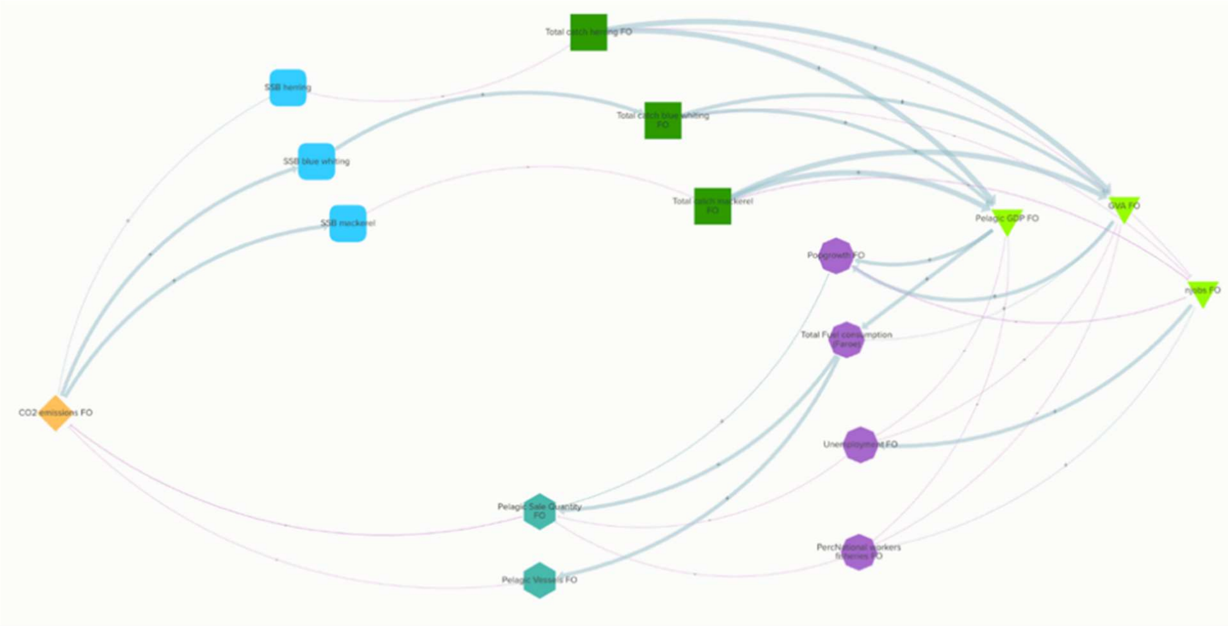

Figure 36. Kumu results for the Faroe Islands – final version.

The high interconnectedness of the SES in the Faroe Islands, as was illustrated in the first version of Kumu results for the Faroe Islands (Figure 32) should be understood as an already simplified version of the SES of the DA. However, as shown through this process, the removal of indicators that do not have theoretical support of relations (Figure 33) and then also indicators for industries that are different from the one in focus (pelagic), resulted in only 39 detected loops, which may not detect the SES complexity, but it did result in more logical feedback loops. For example, in the guidance, it is noted that in assessing loop polarity, an even number of negative links represents a positive feedback loop, such as the one in Figure 36. An odd number then of negative links is likely to be a balancing loop, as in Figure 37. Yet, as also noted in the guidance, it is important to also consider the logic of each loop that is identified to avoid mislabelling.

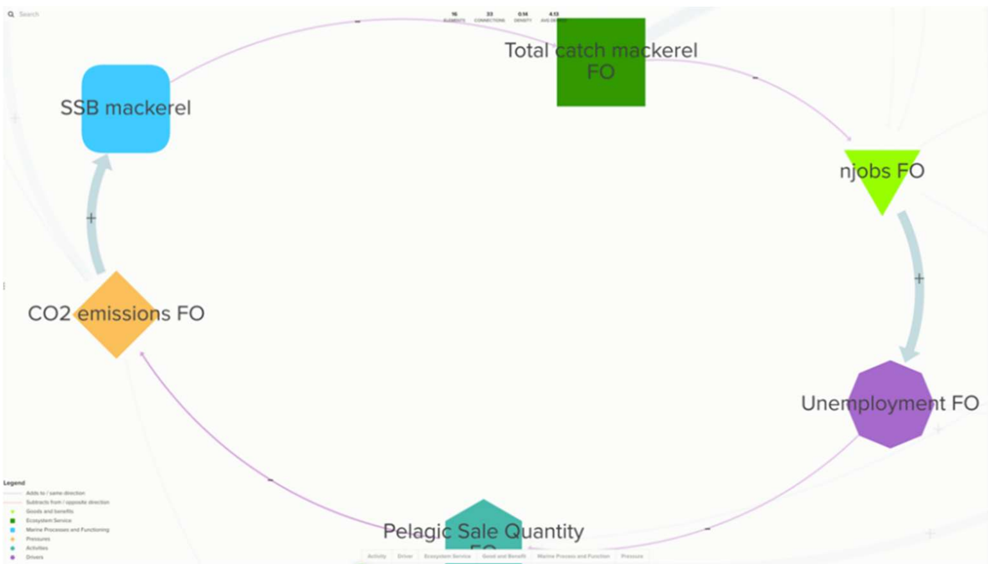

Figure 37. Example of a reinforcing feedback loop in Faroe Islands.

Starting with Figure 37, shows that SSB of mackerel negatively affects mackerel catch, decreasing jobs, which increases the unemployment rate, leading to a decrease in pelagic sale quantity, leading to a decrease in CO_2 emissions and then an increase in SSB mackerel. There are several elements here, that are misleading. Whilst the indicators above do interact with one another, the actual direction of arrows and the strength of the relationship (+/-) are in certain cases mislabelled. For example, the relationship between the number of jobs (in fisheries) and the unemployment rate should in theory move in the opposite (-) direction, not in the same direction (+) as indicated in the loop. That means an increase in jobs should lead to a decrease in the unemployment rate, and vice versa. Another example, CO_2 emissions do not necessarily influence SSB, but SSB may influence CO_2 emissions, as high SSB may indicate that there is more fish, meaning less searching time and shorter time spent at sea. Hence, the direction of the arrow is mislabelled. However, one could argue, that the release of emissions could affect SSB, as high emissions would negatively affect SSB. If applying this argument, the loop should indicate an opposite (-) relationship.

Figure 38 indicates that the SSB of Blue Whiting leads to an increase in total catch, increasing the economic contribution of fisheries, which increases total consumption, leading to an increase in pelagic sale quantity and decreasing CO_2 emissions in fisheries. The logic of this loop represents some of the same concerns as above. Specifically, the direction of the arrows/connections, and the type of relationship, are in some cases mislabelled or misleading. In other words, it is argued here, that all the indicators are connected, but the logic of the loop is not always in place. For instance, as was argued above, the relationship between CO_2 emissions and SSB should affect each other the other way around. In addition, an increase in sale quantity should lead to more emissions, as this means more fishing at sea. Hence, these should have a reinforcing (+) relationship, not opposite (-). However, the loop could also indicate that the fleet is fishing more efficiently, but the reinforcing relationship between total fuel consumption and sale quantity is thereby misleading. Hence, fixing some of the issues raised above concerning the ISA, specifically the matrices, would perhaps overcome some of these concerns about the cause-and-effect relationship between indicators. Second, an issue could also be the preliminary exercise of defining and establishing relevant indicators. Arguably, the guidance allows for multiple ways of interpreting the DAPSIW(R)M elements which come with disadvantages and advantages. The advantage is that it allows a greater opportunity for addressing relevant social and ecological elements in each space. However, it does open the path for misunderstanding and multiple interpretations that make harmonization and comparisons challenging.

Another final thought is that the cause-and-effect between indicators within the same category group is not analysed using the ISA and Kumu, although they could be important or interesting indicators in the SES. For example, in the Faroe Islands, population growth is closely associated with the unemployment rate, as the country has a rather mobile population. To give a specific example, in the 1990s, there was an economic crisis in the Faroe Islands which led to an unemployment rate of 25%, resulting in 10% of the Faroese population emigrating. This, among other things, was caused by poor political decision-making, which also affected the fishing industries. It is arguably a loss to lose connections in the same category (in this case Needs), as they also influence social and economic elements within an SES.

Figure 38. Example of a balancing feedback loop in Faroe Islands.

Overall, the relationship/connection between indicators using Kumu does pose some challenges when making cause-and-effect relationships. The removal of weak or uncertain connections does create more realistic loops, but these loops do not take into account other relationships that are arguably important in an SES. The direction of arrows within the loop and the strength of relationships between indicators deserve more analytical engagement, as the logic of causal loop diagram can be questioned or allow for multiple interpretations.

Kumu results – Greenland

Figures 39 and 40, illustrate the overall results of the Kumu application, both before and after removing connections with no theoretical support or known cause-and-effect relation.

In the Greenlandic case, Kumu detected 432 loops, in the version with reduced indicators (Figure 39). The first impression is that Kumu can be a very useful tool to visually illustrate the complexity of any given social-ecological system and how everything is connected on some level and how indicators affect the system. However, a closer interrogation of the single loops, shows similar problems with the logic of the detected casual loops, as previously described by Iceland and the Faroe Islands. This indicates again that the comparison alone by BOT graph analysis is not sufficient to result in plausible results. The example in Figure 40, which is just one of many examples, illustrates that clearly.

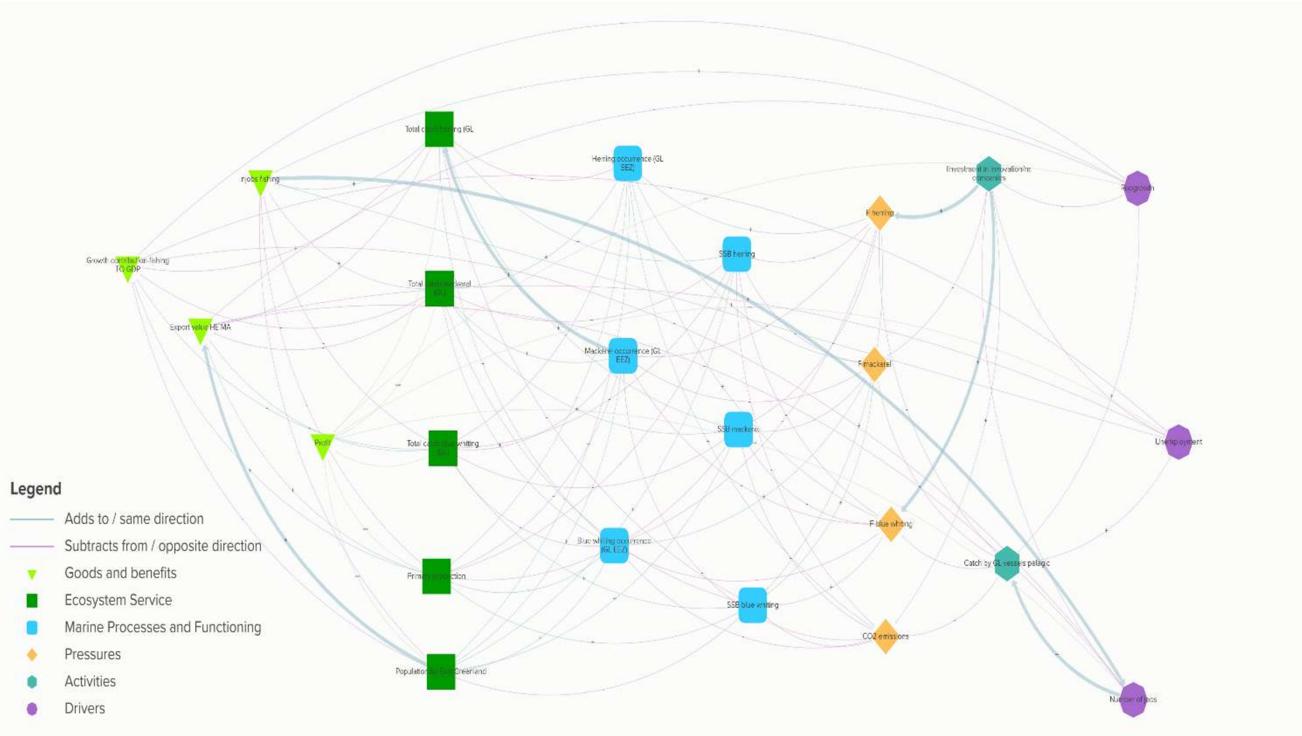

Figure 39. Kumu results for Greenland with all connections included as originally designed, with 24 elements and 100 connections.

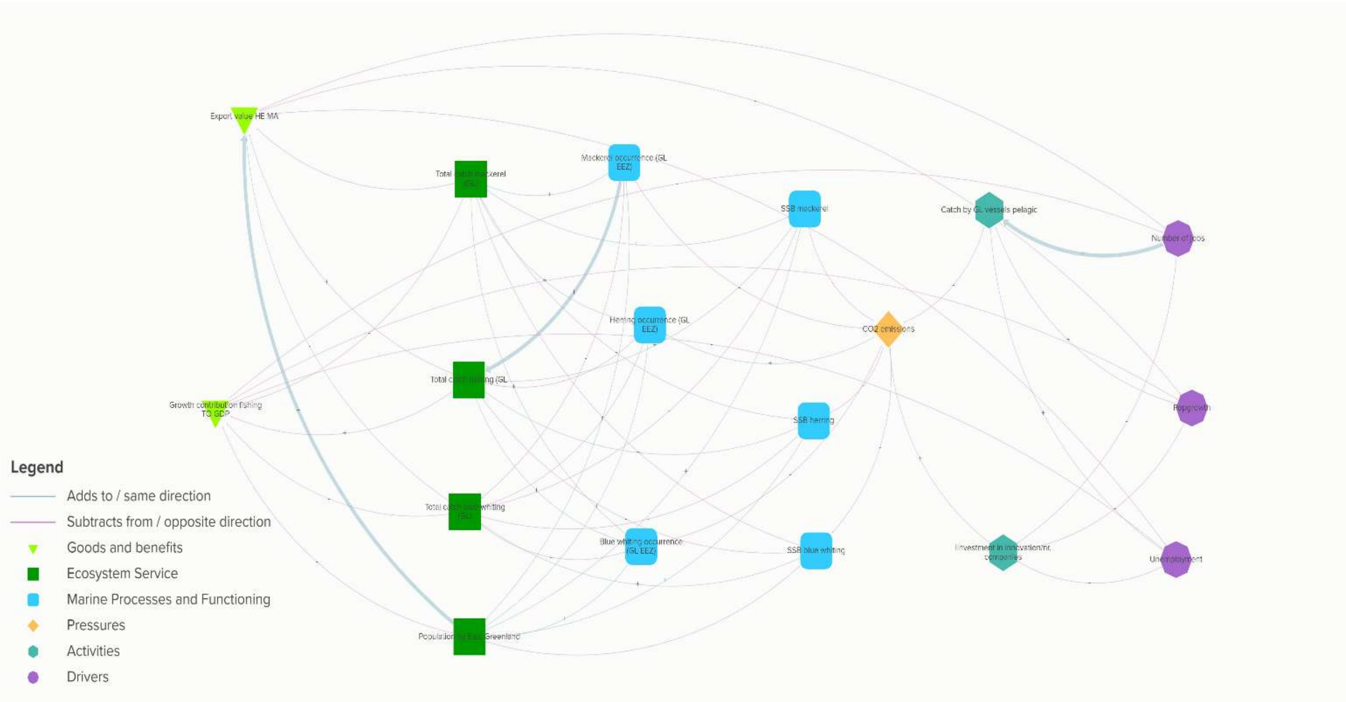

Figure 40. Kumu results for Greenland without connections marked as "1 = no theoretic relationship between indicators" in the matrices shown before with 20 elements and 66 connections. Kumu completely excludes indicators without connections, i.e. that may be important for the whole system, but not necessarily have a connection to other indicators (e.g. fishing pressures and primary production).

Figure 41 suggests that with increasing export values of mackerel and herring, the number of jobs in fisheries will decrease, decreasing the number of companies in fisheries, increasing CO_2 emissions, which again will decrease SSB of mackerel. SSB mackerel will increase the total catch of mackerel, but the total catch of mackerel in the Greenlandic EEZ will decrease the export value of mackerel and herring.

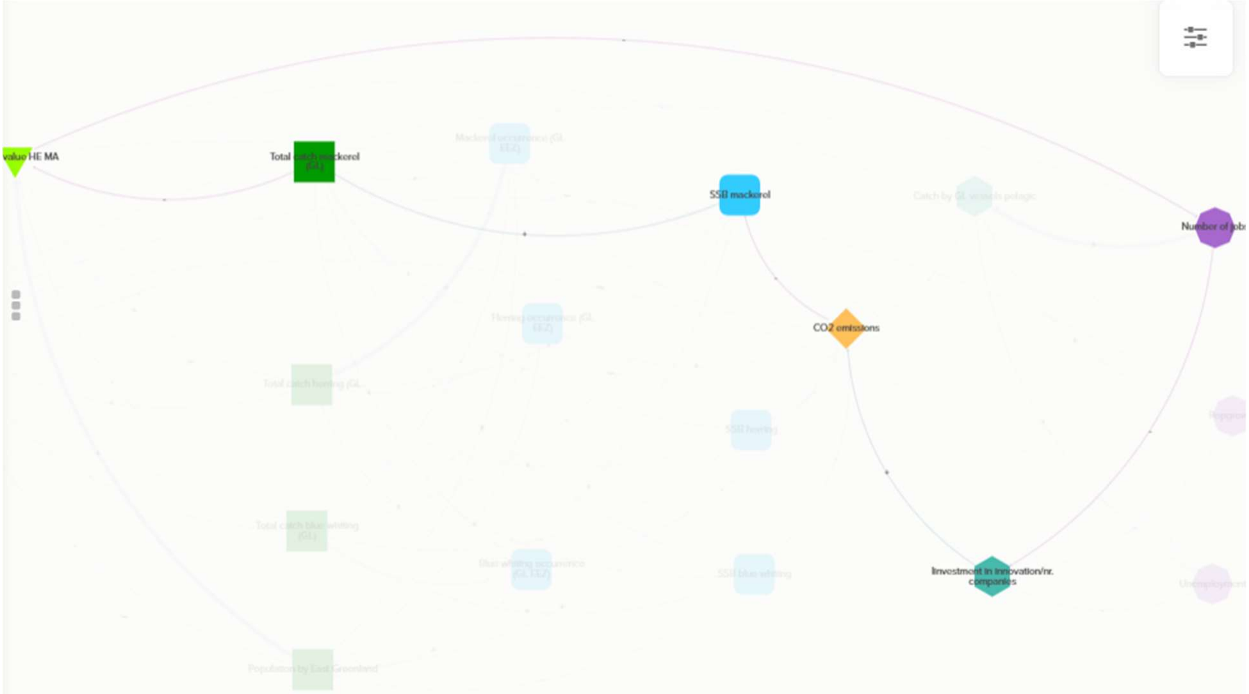

Figure 41. An example of a causal loop, in the Greenlandic Kumu

The casual loop is thus not logical and does not accurately represent feedback between the indicators. Although all indicators are in reality connected, the relationship between indicators, as illustrated in the figure, is not at all represented in theory or even in reality. Thus, the comparison of BOT graphs alone did not result in an uncontested casual-loop analysis using Kumu. As previously discussed, the problem can be connected to the available data used for the indicators. Especially when there were no clear trends in the data sets, with relevant variation over time, the comparison of the BOT graphs between indicators, gets very difficult. This confirms that establishing a relationship between indicators is not always straightforward nor sufficient on its own but requires expert and broad interdisciplinary knowledge of the study area.

As previously discussed, remedying some of the issues raised above concerning the ISA, specifically the matrices, could perhaps overcome some of the concerns about the cause-and-effect relationship between indicators, as well as more time to choose even better indicators. However, there are knowledge gaps and data gaps that will always be a limitation to this method.

The classification of feedback loops as reinforcing (e.g., positive feedback) or balancing (e.g., negative feedback) is a very time-consuming process, as pointed out by both Iceland and the Faroe Islands. There does not seem to be a function in Kumu that can help the process, by automatically identifying these specific types of loops. This would clearly be an important improvement to make the software more user-friendly. From a practical side, it was impossible to find a function that allowed making better screenshots of the single loops, than the above. There is an export function that helped to make Figures 40 and 41, but it does not seem to work for the single loops. This made reading the direction of the arrows quite difficult and is definitely something that makes Kumu user-unfriendly and excessively time-consuming.

Concluding remarks

Overall, this section has aimed to show the process and evaluate the application of the SES guidance. The section has had a focus on navigating through ISA and Kumu from using the DAPSIW(R)M framework to identify indicators to conducting a loop analysis. In this part, a summary and evaluation will be provided, with recommendations for overcoming some of the uncertainties we experienced in applying the guidance.

The starting point of identifying indicators was done separately for each area in the Arctic DA. This decision was made due to the specific social systems of each region within the DA. In terms of ecosystem services and marine processes, the indicators were similar or the same for all areas. It was initially discussed to combine and develop one SES for the DA at a later stage, but since the social systems vary significantly from one region to the other, the DA experts decided to not take this step.

The identification of indicators is a product of interpretation, which may vary significantly from organization to organization, or individual to individual. For instance, Sjókovin (Blue Resource) as an organization conducts research on socioeconomic factors in fisheries and aquaculture, whilst the Marine and Freshwater Research Institute conducts research on marine and freshwater, and WWF Greenland works with the conservation of natural resources. Each organisation may approach the process of identifying indicators differently. In general, much communication between partners revealed the uncertainty within the indicator exercises. Although broadness in defining indicators is advantageous in terms of giving space for local knowledge and expert judgment, it also runs the danger of misinterpretation between the regions. In addition to this, multiple indicators are difficult to quantify in the BOTs over a given time period (which also differed by area due to multiple interpretations), but their importance for the SES in all regions may still be highly significant. The exercises therefore end up excluding crucial SES indicators. Therefore, a suggestion is that the exercises be more defined, and perhaps include some criteria for the process of identifying indicators. Additionally, challenges concerning important indicators that cannot be quantified should be considered to allow a more comprehensive representation of SES. This also raises the issue of the lack of indicators, as a comprehensive visualization requires the integration of a large number of indicators to express the complexity of SES. Another suggestion is to set clear goals and purposes that can serve to inform the very first stage of setting indicators. For instance, considering the indicators related to demersal and aquaculture in the Faroese case – these are important for the SES of the Faroe Islands, but their connection to pelagic species and pelagic fisheries is difficult to assess as these are two separate industries and a decline in one may not affect the other. So overall, there is a lot of uncertainty already in establishing the most relevant indicators, and the more they are narrowed down to only the ones with high certainty, the higher the risk of oversimplifying the system, rather than just simplifying it.

In this perspective it is important to mention having only five indicator options per exercise in the template is not sufficient for some of the exercises. The Arctic DA considered three major shared pelagic stocks that cannot be considered as one unit since each is influenced by different factors and has different biological trends. Additionally, each region within the Arctic DA had different accessibility options throughout the last two decades, not only due to ecological reasons but also because of different access rights. This easily increases the number of indicators for the ecological exercises and hence the whole template must be adapted. Due to the interconnectedness between the templates, errors can easily be introduced. Checking for potential errors turned out to be an extremely time-consuming activity.

Completing the matrices represented multiple challenges that affected the whole analysis. First, the BOTs at most times did not have theoretical support. It would therefore be an option to identify indicators using a theoretic approach only. Second, the connections between indicators are almost impossible to measure using the BOTs. This could perhaps be overcome by conducting a time-series analysis or other forms of analysis that allow expertise about indicator correlations to be included. Third, assessing the strength of connections is often difficult to assess using the BOTs and increases the risk of misleading interpretations. As a result, the loops detected in the Kumu application result in misleading and/or simply wrong feedback loops that would be highly problematic to represent to stakeholders. This also raises the issue of the expertise of the individual choosing indicators and assessing relationships. For instance, if an individual does not have a comprehensive understanding of the SES in each region, then they could include improper/unrelated indicators, blindly follow the BOTs, and present these results as a given to stakeholders. This can then result in misleading policy recommendations that can have complications in the real world. In addition, considering that we reduced indicators significantly in Kumu, we are losing a lot of important details and features of any SES, but gaining more logical loops. This then becomes a matter of assessing whether the costs outweigh the benefits when designing a sSES.

Overall, using Kumu requires a high degree of confidence and expertise in the data registered in the Excel sheets. Therefore, it is highly recommended to minimize uncertainty before representing and/or recommending policy options and management methods to stakeholders.

8 Results from the validation of the simple SES model

These results correspond to the methods described in Section 5. The final aim of evaluating the presumed utility of this process is learning from the experience of applying the sSES to the three DAs. As the present study is focused on providing the material basis for improvement of the tool under study, two sets of evaluation were asked from the modelers: first a numerical statement of the degree of satisfaction they had with that specific aspect (criterion or indicator of the method proposed), and second, to justify the numerical statement with comments that complement the value statement. The idea is to understand the answers received from the modelers (henceforward "users"), always with the background goal of providing substance for the developers of the method (i.e., the group that created the sSES) on where, and how to improve it. It is also important to highlight the present process of evaluation was developed with the idea of being applied to a broad set of qualitative models, and therefore some indicators might not apply to the specific case. One assumption of this protocol was to have a different group of stakeholders, other than the users, directly involved in the modelling process. In the cases evaluated here, the model was conducted by a group of specialists (scientists), which is not a problem but limits the reach of some indicators that were previously designed for a non-specialist stakeholder group, such as 6 (Trustworthiness or Guru status of the system dynamicist), 7 (Meaningfulness of the process), 17 (Learning), and 18 (engagement) (Table 32).

The overview evaluation of the process of modelling (Figure 42) was positive (overall average of 3.5 on a range from 1 to 5). The result is due to the low number of individual evaluations of "very dissatisfied" (1) and "moderately dissatisfied" (13), which reveals a low (18%) dissatisfaction with the process against "moderately satisfied" (24) and "very satisfied" (7), revealing the overall satisfaction (40%) with the process. Complimentary, 15 answers were "not apply" (19%), and 18 were neutral ("neither satisfied nor dissatisfied") (23%), summing the 78 evaluations. Nonetheless, these individual evaluations rarely agree on the same criteria, reflecting the individual variations, and therefore the consistency of each quantitative statement must take this variation and the specific comments into consideration. Nonetheless, an average and mode (when possible) were provided as a reference for the final evaluation. These and the other comments therefore can be addressed in the revision of the sSES approach and its guidance as planned in the project description, in Task 3.3.

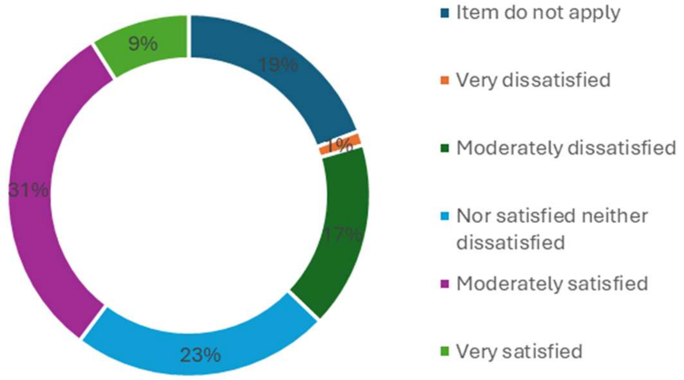

Figure 42. Aggregate distribution of answers from all indicators

Table 32. Tests for presumed utility in qualitative models. A, B, and C are the Demonstration Areas to which this questionnaire was applied. The scale is from 0 - item does not apply (white), 1 – very dissatisfied (red), 2 – moderately dissatisfied (yellow), 3 – nor satisfied neither dissatisfied (grey), 4 - moderately satisfied (blue), and 5 – very satisfied (green). The colours are illustrative of the values. Avg is the average of those results which excludes zero. Mo is the mode, when possible. Every comment was numbered to be referred to in the text.

Type	N	Criteria	Description	A	B	C	Avg	Mo	Comments from users:
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	5	4	3.7		1. Whilst the purpose of simple SES is clear, the actual model or design of the model is not sufficient in reaching the purpose. 2. The purpose is clear and explained in the presentation. 3. This was only reached after some discussion between the areas that compose the DA.
	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?	2	3	4	3		4. Already hard to use for modelers 5. Not user friendly 6. Not usable for stakeholders 7. Not compatible with users capacities 8. Rather academic, so people operating it would have to have academic background. 9. Understanding the network might imply a basic knowledge of the biological system 10. This type of graphic representation can be difficult to interpret and not very immediate to some end users 11. The idea is for the model to be operated by third parties. It's available in an adequate format, but it could be clearer. It could be more compatible, but because it has many variables, it can sometimes be a bit confusing.
	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 modelling 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?	1	4	4	3	4	12. Tables and Data not user friendly 13. Language appropriately set by modellers 14. Loops or big CLD are still too complicated, signs are too confusing (+ for same trend, - for opposite trend) 15. Hard to interpret for stakeholders 16. The use of different colors and sizes makes it easier to understand. I believe that the individual loops are easier and more immediate to understand for users 17. The use of the results can be partially useful, although the indirect implications may be difficult to interpret.
	4	Perspectives in Boundary-adequacy	Refers to different perspectives of issues and policies. Does the modelling 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; d) Policies for doing so?	3	4	4	3.7	4	18. Could be used well to inform, but would need to adapt rating of connections as we suggested in the deliverable using theoretic knowledge instead of BOT diagram comparison 19. a) No; b) Yes; c) Yes; d) No

	5	Norms/values in boundary adequacy	Refers to different perspectives of values and acceptability. Do the models support debate concerning and represent 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?)	2	3	5	3.3		20. It can be argued that the quantitative nature of the model makes it difficult to represent values and acceptability between relevant actors, as many indicators were simply not quantifiable. 21. a) Yes; b) Yes
	6	Trustworthiness or Guru status of the system dynamicist	An affinity with the modeler can enhance positively the modelling process and the Policy Insights or Recommendations (PIoR) implementation. Is it possible to report a positive relationship or atmosphere between the stakeholders and the modelling team?	4	4	5	4.3	4	22. Stakeholder involvement went well overall 23. Yes
	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?	0	5	4	4.5		24. We haven't shown this to stakeholders, not sure what to say here 25. The model was built based on relevant results from stakeholders surveys from WP2.
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?	3	4	3	3.3	3	26. Indicators are mostly representative 27. Some major indicators are missing though, because they would not have any connection with other indicators, but still be very important overall 28. The connections as they are now (being Solely based on the BOTs) do not represent causation in most cases 29. The model allows the essential factors and elements of each system to be included, however the established connections (both direct and indirect) are often "context dependent" and cannot be generalized into a single model. 30. Yes, after we manually modified some of the connections. Connections follow a logic based on current knowledge of the system but do not represent causation as appropriate tests are not always available.
	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?	0	3	0	3	0	31. The analysis of loops can be very long. It is not clear how the most relevant loops should be identified. 32. Loops are useful, but I'm afraid they simplify the complexity of the system a lot, and so the result may depend on who handles the analysis 33. The relevant processes and factors are contextual and not generalizable to the scale of the archipelago. In summary: there is a disagreement between the scale at which this methodology is applied and the scale at which the processes occur or have been measured.
	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?	3	4	4	3.7	4	34. Some indicators are not included in the model, because they would not show any connection or have enough relevant data to be included, but they may still be major contributors and of major importance to the stakeholders 35. The most relevant variables are explicitly represented 36. Variables are represented correctly. 37. Yes, relevant variables are well represented.
	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	2	4	3	3		38. Generalising the model is not possible at this stage of the simple SES. The fuzziness is rather high and the level of uncertainty in identified connections makes it not a useful tool for policy/recommendations.

			model to a similar system with minor adequations? Would it still be meaningful and useful for the creation of policy insight or recommendations?						39. The model is moderately specific, so it will be hard to apply it to other situations with just minor changes.
	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?	2	3	4	3		40. Not necessarily, as the connections themselves are mostly not representative? 41. It is not clear how this should be done 42. The model is not very reliable for making interpolations and is even more so for making extrapolations. 43. Yes
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?	2	5	4	3.7		44. With the status as it is now, it is not recommendable to forward this or use it as policy insight or recommendation tool 45. Yes, the model highlighted activities in the model with the greatest number of connections suggesting they play a key role in the system and should be prioritized when considering management plans and/or recommendations.
	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?	0	4	0	4	0	46. With the status as it is now, it is not recommendable to forward this or use it as policy insight or recommendation tool 47. Identified PIoRs rarely identify recommendations that are not previously known.
	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 modelling process?	0	4	0	4	0	48. With the status as it is now, it is not recommendable to forward this or use it as policy insight or recommendation tool
	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?	3	4	0	3.5		49. Not sure how to rate this here, as the scale is satisfied to dissatisfied and not agree/don't agree… 50. The PIoR would require change if applied to a different location as indicators may be different/ other indicators may be more important
	17	Learning	Do participants state that they learned about the system, other stakeholders (the community), or the policy-making process during this modelling process? Are they satisfied with that? If they want to learn more, did they receive support on how to do that?	0	4	4	4	4	51. Yes, during the process, participants discussed and brainstormed about how the different activities impacted or were impacted by the indicators, which in turn revealed to be a learning process.
	18	Engagement	Did stakeholders engage in any group/action related to the issues dealt with in the modelling, during/after the modelling exercise?	0	5	0	5	0	
Administrative, review, and	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?	2	3	5	3.3		52. It is quite time consuming and not so easy to add/edit new information to the model and update the results 53. New or updated information can be added to the model.
	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 modelling process concluded within the expectations of time and costs? Are there recommendations to improve the efficiency of the modelling exercise for the next team or exercise?	3	3	3	3	3	54. It would definitely be recommended to improve the efficiency of the modelling exercise for the next team or stakeholders 55. The main issue was that we took more time to collect some data (i.e we had to contact the data owners several times to obtain the data)

	21	Documentation	Refers to the adequacy of the process of making every step in the modelling process replicable by taking a formal process or writing assumptions, discussions, updates, or a change in previous steps regarding the modelling process. Is the model satisfactorily documented?	4	3	4	3.7	4	56. The process was well-documented in Deliverable 4.1 57. Yes
	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?	4	2	3	3		58. This is not easily reproducible as the access to the underlying data may be an issue and the understanding of other potential users of the data and sources may differ substantially as well. However, the process was documented. 59. Many doubts about the results, given that there is not a formal procedure on how to conduct network analysis, identification of most relevant elements and loops, how to use the CLD to predict system behaviour in extreme conditions, etc. 60. We gave special focus to the connections that were not recognized immediately by the model. According to the data alone, some connections were identified as negatives or positives, but they did not make sense or did not conform with scientific or expert knowledge. In these cases, we manually overridden these connections.
	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 modelling 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?	2	4	0	3		61. Since the model is not useful at this stage for PIoR, it would not pass an audit or cross-validation
	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, modelling 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 modelling exercise?	2	3	0	2.5		62. Since the model is not useful at this stage for PIoR, it would not pass an audit or cross-validation
	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?	2	3	0	2.5		63. Since the model is not useful at this stage for PIoR, it would not pass an audit or cross-validation
	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 modelling exercise? Are these changes congruent with the desired state modeled?	2	2	0	2	2	64. Since the model is not useful at this stage for PIoR, it would not pass an audit or cross-validation 65. It is not clear how this should be verified

In addition, the four types of indicators (Guidelines and Processes, Specific Model Tests, Policy Insights and Spillovers, and Administrative, Review, and Overview) present different mosaics of satisfaction and dissatisfaction that point to different aspects and steps of the modelling process. This division into four dimensions of the validation protocol represents distinct elements in the modelling process that allow a nuanced analysis of the process from the start (guidance and purpose); of the represented indicators, model assumptions, boundaries, variables, etc. (specific model test); an evaluation of the capacity of the model to create some spillover or change in the system (policy insights and spillovers), and, to its documentation and replicability (administrative, review, and overview).

The distribution of results from the evaluation of each dimension (Figure 43) shows that the previous steps of the modelling process (guidelines and process) had more positive evaluations. These positive valuations start to decline along the progression to the other dimensions, where they are substituted by the increase of "not apply" or "neutral" evaluations. These results were expected since some of the later indicators (here taken as synonym of criterion from Table 1 and 32) represent challenges to the modelling process and rarely would have a positive evaluation if the process is at its early stages of development. Nonetheless, despite this expected trend, the evaluation can be profitable for the developers as it provides some insights into the performance of the modelling process and pinpoints where some improvements might be made. The overview result from the four dimensions brings some insights (discussed later in detail), such as:

1. An overall positive result of Guidelines and process dimension;
2. Why does the Specific model test dimension have so many "neutrals", while still being positive?
3. Why did "policy insights and spillovers" have this mix of "not apply" and "positive" evaluations?
4. What happened in the Administrative, review, and overview dimension that shows a dominance of "neutral", "moderately dissatisfied" and "not apply" results?

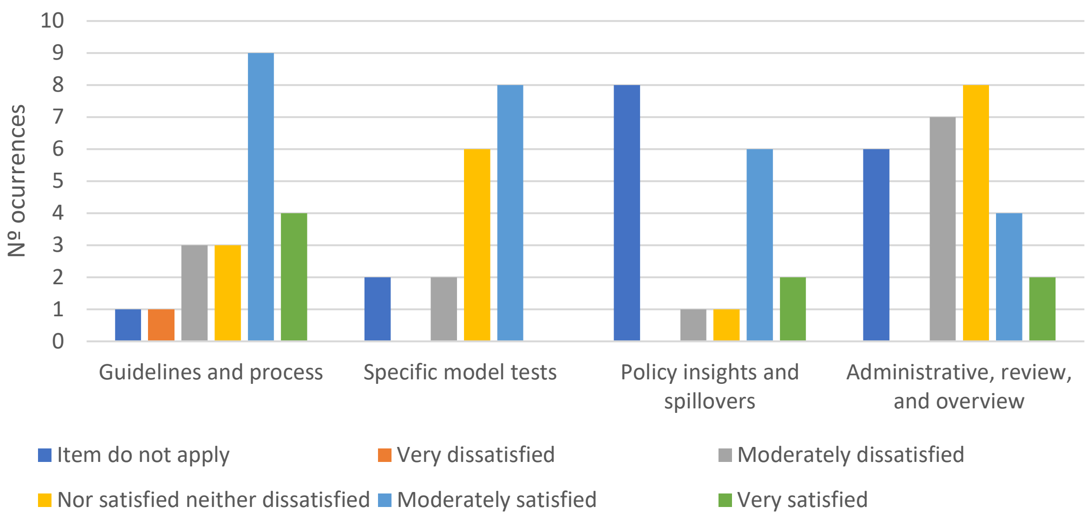

Figure 43. Aggregate distribution of answers (all indicators)

To understand these insights, an analysis of the quantitative evaluation and the comments provided for each indicator will be used. The modelers provided 65 comments referring to these indicators and the whole process (Table 32), these comments were numbered and will be referred to in each indicator description. A word cloud was provided (Figure 44) to illustrate the comments and highlight words such as model, connections, indicators, relevant, data, loops, useful, etc. While this is of limited scientific value, can bring an appetizer of the main topics covered by the comments.

Figure 44. word cloud made with the comments (words appeared three times or more).

8.1 Dimension 1: Guidelines and process

This dimension embraces the indicators 1 to 7 (Table 32). The more positive aspects found in the "guidelines and process dimension" (meaning criteria where the set of evaluation results are made from "all positive" or "positive with neutral" or "positive with not apply" and no unsatisfactory result is present in that indicator) were in indicators 4 (Perspectives in Boundary-adequacy), 6 (Trustworthiness or Guru status of the system dynamicist) and 7 (Meaningfulness of the process). Indicators 1 (Purpose), 2 (Usefulness), 3 (Presentation), and 5 (Norms/values in boundary adequacy) had a mixture of "positive, neutral, and negative" evaluations.

Criterion 1 (Purpose) (average 3.7) had two positive evaluations and one negative. The positive allows us to understand researchers considered the purpose of the whole exercise "satisfactorily" and "very satisfactorily" stated a priori (comments 1 and 2). The "moderately dissatisfied" answer indicates the contrary but comment 1 seems to point out that the problem was not the purpose but the current state of the model and its capacity to reach its goals. In addition, some users said the understanding of the goal was reached after deliberation of the users (comment 3), which reveals that the step of initial group deliberation is important to the process as an introductory step.

Criterion 2 (Usefulness) (average 3) brings a conflicting message from the comments with the evaluation. Despite having one “moderately satisfied” value (supported by comment 11), all other comments (4 to 10) say the contrary. As the other two evaluations were “moderately dissatisfied”, and “neutral”, these comments might represent some important traits of the model that need to be improved. Comments (4-10) question the adequacy of the model to be used, communicated, or understood by stakeholders, which is a relevant topic to be considered regarding its usability and communication capacities when approaching a larger stakeholder audience.

Criterion 3 (Presentation) (average 3, mode 4) is curious because brings two “moderately satisfied” values associated with the only “very dissatisfied statement” in the whole assessment, revealing the opinions are very divergent for this indicator. Here the idea was to understand if the model and materials used along the process are formatted according to the audience's educational level (a proxy for the level of scientific knowledge), and also the clarity of the CLD. As it is notorious that CLDs can be very confusing if they have too many variables, loops are not made clear or the connections are tangled, this indicator is of great importance. In the present case, comments are also mixed. Some comments (12 to 15) show that the tables, the data, the language, and the CLD are complicated to interpret. In addition, the implications (comment 17) of the modelling may be difficult to interpret. On the positive side, the use of different colours (comment 16) can be useful in facilitating understanding.

Criterion 4 (Perspectives in Boundary-adequacy) (average of 3.7 and mode 4) reveals the capacity of the process of modelling to support the debate on different issues of the system, especially the “issues to be addressed”, and the “goals to be achieved”, which were positively evaluated (comment 19). Some limitations were appointed (comment 18) regarding the rating [sic] of connections (what we assume is the link-polarity or link intensity due to the extensive discussion of this topic in section 7).

Criterion 5 (Norms/values in boundary adequacy) (average 3.3) brings diverging evaluations with one “moderately dissatisfied”, one “neutral” and one “very satisfied”. Comment (18) restates that some variables were excluded because there were no metrics for them, despite their relevance, and also claims difficulty in representing values and acceptability in a “numerical” model. Here the misunderstanding of what is a CLD and what is a numerical model seems relevant. This might reflect the idea that the data selection was made mostly focusing on variables that have an associated metric, and discarding the other variables that would have been important, but that were not supported by an obvious quantitative indicator, which is problematic. This problem is recurrent in many indicators and points out a lack of understanding of the modelling process by the users. The other comment (21) states a confident “yes” to the questions regarding the acceptability of the goals and policies (by the audience) but as this model was not presented to the stakeholders (the presumed audience), it indicates this acceptability is probably inferred from the opinion of the modelers.

Criterion 6 (Trustworthiness or Guru status of the system dynamicist) (average of 4.3 and mode 4) shows that, in the view of the modelers, the stakeholders of this process (consulted previously, not during the modelling process) had an affinity with them (comments 22 and 23), with no complaints of the interactions of the modelling team with the stakeholders they represented.

Criterion 7 (Meaningfulness of the process) (average 4.5) As the result from this indicator includes one evaluation of "not apply" (they have not shown the results to stakeholders and therefore they are not sure of how to answer this indicator – comment 24), the average includes only two positive evaluations. The other comment (25) implies the stakeholders were considered in the modelling process (from previous consultation), which was considered positive for the modelers. As this indicator was trying to capture stakeholder experience in the process, the reach of these results is limited as the modelling was made by a group of specialists.

Overall performance in the Guidelines and process dimension was positive and the topics of interest are:

- The relevance of an initial group deliberation as an introductory step, that was not satisfactorily done in the present case.
- The question of the inadequacy of the model to be used, communicated, or understood by stakeholders.
- The tables, the data, the language, and the CLD are complicated to interpret.
- The limitations of connections (probably the link-polarity or link intensity attribution).
- The selection of variables was made mostly focusing on variables that have an associated metric, despite those considered relevant.
- Stakeholders showed a great affinity with the research group.
- Stakeholder experience in the process was limited due to their absence during the model phase.

8.2 Dimension 2: Specific model tests

This dimension embraces indicators 8 to 12 (Table 32). The more positive indicators in this dimension were 8 (Structure verification) and 10 (Boundary adequacy (as structure)). Indicator 9 (Loop Polarity) had no positive or negative evaluations, and the others (indicators 11 and 12) had a mix of positive and negative evaluations.

Criterion 8 (Structure-verification) (Average 3.3 and mode 3) This indicator was considered positive by only one evaluation, and neutral by the other two. This indicator is pointing at the core of the model results, which in a CLD is its structure, therefore comments are important. Some of these comments need clarification (e.g., 26 - indicators are mostly representative!). Others are clear and relevant, such as "Some major indicators are missing though because they would not have any connection with other indicators, but still be very important overall" (comment 27). Why was a major indicator not included? As this also happened in other DAs (as described in section 7), there is an understanding that the users excluded some relevant indicators due to: a) lack of data, b) problems on how to connect them to the other elements of the CLD, c) incongruence with expert knowledge, d) different time/spatial frames e) lack of computational capacity, f) to avoid data gaps g) incapacity for connecting indicators at the same level in the simple SES framework. Part of this problem suggests a mismatch between the scope of what modelers and/or stakeholders believe to be important socially or ecologically, versus the scope of what is possible to include in the model while maintaining credibility. That is, if the importance of an element of the system is extremely high, it is difficult to include it if its relationship to other elements is also highly uncertain. As this step is crucial to dictate the structure of the system, it is recommended that the developers clarify how to select and use indicators in the modelling process, as well as suggest ways of dealing with uncertainty.

In addition, comments such as “The connections as they are now (being Solely based on the BOTs) do not represent causation in most cases” (comment 28) and “The model allows the essential factors and elements of each system to be included, however, the established connections (both direct and indirect) are often ´context-dependent´ and cannot be generalized into a single model.” (comment 29) indicates how some users are connecting variables based on their BOT, and not based on causality (from previous understanding of the system). This is crucial as connecting variables from BOT might be misleading (such as establishing spurious correlation). This problem is acknowledged by comments such as ”Yes”, after we manually modified some of the connections. Connections follow a logic based on current knowledge of the system but do not represent causation as appropriate tests are not always available.” (comment 30).

Criterion 9 (Loop Polarity) (average 3 and mode 0). This indicator compares the loop polarity in the model with the previous knowledge of the user. Here the comments focus on different parts of the modelling process to be adjusted. Regarding the loop analysis (comment 31) states that the loop analysis is very time-consuming and it is not clear which loops to use. As this is congruent with many other feedbacks from the users (during meetings, emails, and section 7), it is clear that in-depth orientation on how to produce, select, and use feedback loops is fundamental. The issue of how to select the relevant ones is restated by comment 32. In addition, as some groups declared they had identified more than 500 loops (by email), probably the result of the automation of this part of the exercise surpasses the human capacity to deal with the results, and therefore the final application of this exercise might lose its meaning. It is also relevant that all three DAs revealed they reduced their structural complexity to fit into the Kumu software capacities.

Comment 33 points to a disagreement between the scale of the model and the scale where the processes occur in the DA. This comment is also corroborated by a declaration of the multi-level challenges for governance by some DAs (Section 7). From this, it can be interesting for the developers to investigate how the delimitation of the study areas was done (see Sections 7 and 9). As this project used an innovative spatial delimitation of the DAs, some groups questioned how to embrace the diversity of data, governance systems, representativeness of data, different spatial and time frames of the data, etc. Furthermore, there was a divergent attempt how to embrace data diversity: Macaronesia did an average of the data or a "best guess" when possible. Tuscany went for an "overarching perspective of the archipelago", with a "best guess" approach for the data and the Arctic chose not to integrate the data as the diversity of the social structures in their SES was so divergent, which impeded the process of transforming the study in one representative case, despite the stated "great similarity of the ecological realm" (all from section 7). Boundaries can be treated as spatial delimitations of the study, but also in terms of governance, representativeness in the modelling process, values and perspectives regarding the system, or others that seem of interest to match with the overall goals of the modelling exercise.

Criterion 10 (Boundary adequacy (as structure)) (Average 3.7 and mode 4). This indicator had two positive evaluations from two users and one "neutral" from the third. This indicator tries to capture if the model represents the relevant structures in the system. Most of the comments (35 to 37) answered positively to this question affirming the model does represent adequately the variables, which seems contradictory with the statement in indicators 5 (Norms/values in boundary adequacy), 8 (Structure-verification), and 9 (Loop Polarity). In addition, there is a comment (34) which restates the exclusion of indicators due to points listed in the indicator 8 description. As it sounds more coherent to understand the users had some issues with the inclusion/exclusion of variables, this indicator might be useful to have a broader view of the topic, as suggested previously.

Criterion 11 (Family-member) (Average 3) This indicator tries to embrace the capacity of the model, with few adjustments, to represent a general case, instead of just the case it was built for. In this case, the results are a mixture of positive and negative evaluations. Comment 38 reveals the low confidence in the current state of the model to guide policy-making and generalizations. As this might change with the latter development of the model, it can be interesting to understand the "fuzziness" and the "high level of uncertainty in connections" as points of interest. More broadly, comment 39 reveals that the model is "moderately specific" and finds it difficult to apply in other situations with minor changes.

Criterion 12 (Extreme-conditions) (average 3). As this variable tries to capture some possible extreme behavior (usually associated with numerical models), its applicability in the present case is controversial. With both positive, negative, and neutral evaluations, there seems to be no consensus on the satisfaction of its applicability as well. Nonetheless, comments question the applicability of this indicator (comments 40 and 41) and how to do this analysis. The authors agree this indicator can be more appropriate to a numerical model or to a later stage CLD, where an attempt to capture possible extrapolation of the extreme results would make more sense.

Overall performance in the Specific model test dimension was a mixture of positive, and neutral evaluations (with few negatives). The many neutrals show some topics might have prevented a more positive evaluation of this dimension. The topics of interest are:

- Users excluded relevant indicators due to: a) lack of data, b) problems on how to connect them to the other elements of the CLD, c) incongruence with expert knowledge, d) different time/spatial frames e) lack of computational capacity, f) to avoid data gaps g) incapacity for connecting indicators at the same level in the simple SES framework.

- Some users connected variables based on their BOT trends, and not based on causality.
- Loop analysis is very time-consuming; there are too many loops to consider; users reduced their number of variables to fit Kumu´s capacity.
- The users need guidance on how to establish geographical boundaries and on the integration of different data sets, governance systems, goals, etc. Guidance on how to create an "overall representative DA" is necessary for consistency.
- The model is "moderately specific" and probably difficult to be applied in other situations with minor changes.

8.3 Dimension 3: Policy insights and spillovers

This dimension embraces indicators 13 to 18 (Table 32). All indicators in this dimension, except indicator 13, were positive.

Criterion 13 (Insight generation capacity) (average 3,7) embraced the idea that the modelling process might have already produced some policy insight or recommendation (PIoR). As this indicator had one answer "moderately dissatisfied", and two positives (one "moderately satisfied" and one "very satisfied") it was expected that some policy insight was already present in the comments or discussions, but the comments don´t support this affirmation. Comment 44 states against recommendation at this point of development (more development needs to be done), before considering the PIoR. Comment 45 suggested the activities in the model with a high number of connections might be of interest to policy development, which despite not referring to a policy de facto, is a positive output of the process due to the identification of a target for a PIoR to be developed. If this is correct, then we have an inconsistency with indicator 14 which said no novel knowledge was found.

Criterion 14 (Relevance and Fertility of PIoR) (average 4, mode 0) tries to understand the usefulness of the model in bringing new recommendations to manage the system. As the results brought two answers "not apply" and one "moderately satisfied" the results should be seen with caution. The comments in this section point out that with the current state of development, no policy recommendations can be taken (comment 46) and the results rarely point to recommendations that are not previously known (comment 47). As the model might reach a higher level of maturity, we infer that its capacity to generate policy insights will also grow. On the other hand, as many comments (in this evaluation and in section 7) pointed out that some links disagree with expert knowledge, we call attention to the fact that this tool might not corroborate previous knowledge and as seen, complement it with additional relevant insights. To understand this, the statement that "recommendations are not bringing anything new" (comment 47), might be of interest.

Criterion 15 (Congruence of PIoR with culture) (average 4 mode 0). As this indicator embraces the possible acceptability of the policy insights by the community, it would make more sense when the results were presented to a wider audience. Nonetheless, some users found it useful already and attributed a "moderately satisfied" with what they have so far.

The other two users attributed a "not apply" to this indicator, justified by comment 48. The authors agree that at this step it is difficult to see the acceptability of the recommendations as they are not fully developed. This result can be better treated in the future with more information.

Criterion 16 (Boundary adequacy (as policy)) brought a mixed result of one answer "neutral", one "moderately satisfied" and one "not apply". As this criterion was relative to understanding if the policies and recommendations proposed here would work elsewhere, the results are dependent on indicators 13, 14, and 15. One comment (49) stated it was unclear how to answer this question. The other comment (50) pointed out that to apply for a different location the policy insights would change as other indicators might be more relevant. This is also a reflection of indicator 11 which tried to understand the spillover effect of what was learned here to broader or different areas. From these indicators, it is understood that the results are more local than transferable. They can be potentially transferable with new efforts in modelling and data acquisition.

Criterion 17 (Learning) (average 4 and mode 4). The idea here was to understand if people involved in the modelling process learned from the experience and how they evaluate that learning. And the answers were a "not apply" with two "moderately satisfied". The only comment on this topic (51) pointed out the participants had to think about "how activities impacted or were impacted by the indicators", and that was evaluated positively. It is important to see the users evaluating positively the learning process they went to because that is one of the main goals of any modelling exercise. Nonetheless, this indicator would make more sense when applied to a group of people with different knowledge backgrounds, and different scientific training levels, where complementarity of information and knowledge could evolve in discussions and mutual learning.

Criterion 18 (Engagement) (average 5 and mode 0). This indicator tried to capture if stakeholders already started to change their system after/during the learning they had in this process. As there were no comments here, the understanding of two "not apply" vs one "very satisfied" is compromised.

Considering the model was built by groups of scientists and none reported any engagement action from the stakeholders, the proper answer here would be there has been no engagement action so far, but one of the teams sees great potential for engagement in the future (which would make sense of the answer "very satisfied").

Overall performance in the Policy insights and spillovers dimension was positive with many "not apply" that reveal some steps in the modelling process might have not reached the necessary maturity yet. Some topics of interest are:

- Targets for policy development were found (this statement is incongruent with other indicators).
- Results rarely point to recommendations that are not previously known (low confidence in this statement).
- Users indicate results can be potentially transferable with new efforts in modelling

and data acquisition.
- Users considered positively the learning they had along the process.

8.4 Dimension 4: Administrative, review, and overview

This dimension embraces indicators 19 to 26 (Table 32). Two of the indicators in this section were strictly positive (19, and 21). One indicator had three "neutral" evaluations (indicator 20). The other five indicators were a mixture of positive with negative evaluations (22, 23, 24, 25, and 26).

Criterion 19 (Ease of Enrichment) (average 3.3). This indicator embraces the capacity of the model to be enriched by new data or other assumptions. As the results were one "moderately dissatisfied", one "neutral", and the other "very satisfied", the evaluation is multifaceted. Comment 52 shows that updating or adding new information is time-consuming. This is coherent with the descriptions made before in this section. Comment 53 states it is possible to input new information, without giving a judgment of how easy this process is, which was the topic of this indicator. Therefore, considering the many statements about the time-consuming effort to translate the information from the tables to the CLDs in Kumu (see also comments 31, 54, and 55), it is a point of interest.

Criterion 20 (Time & Cost of the Intervention) (average 3, mode 3). Every group agrees this is a "neutral" evaluation. The comments converge that the efficiency of the modelling exercise could be improved (comment 54) and that they took an undesired time to collect the data (comment 55). This is congruent with comments in section 7 where some users complained about the time requirement for dealing with the tool (page 76), for attributing loop polarity (see also Indicator 9), and others.

Criterion 21 (Documentation) (average 3.7, mode 4). This indicator was focused on the capacity of registering the steps taken and the data consulted in such a way someone in the future could re-access it. For this indicator, two users evaluated it as "moderately satisfied" and one as "neutral". As the comments agree that the documentation was well done in the present deliverable, and there were no statements otherwise, it is understood the users are positive about the level of documentation of the process.

Criterion 22 (Replicability) (average 3) Embraced the belief that someone else could reproduce the mode based on the documentation (indicator 20). The results for this indicator presented one "moderately dissatisfied", one "neutral", and the other "moderately satisfied". The three comments questioned the capacity for reproduction of the model not due to the documentation (see comment 58), but because of the need to access information that is not documented, such as data (comment 58), the absence of a formal procedure to conduct a network analysis, to identify the most relevant elements and loops was also appointed as a limitation to reproduce the results. The topic of link polarity also came back (comment 60) as this was seen as an "overriding command" from the expert group that can be troublesome for the person interested in replication as it probably is poorly documented.

Criterion 23 (audit or cross-validation) (average 3). This indicator tried to embrace a new look at the model, testing if its assumptions and structure make sense. We had conflicting results, one "moderately dissatisfied", one "not apply", and one "moderately satisfied". The only comment points out that the mode is not useful at this stage (comment 61). Considering the many comments before regarding the contradictions between what the model presented and the expert knowledge, this indicator might require attention from the developers in terms of coherence of the loops, link polarities, and conflicts with expert knowledge.

Criterion 24 (higher-level model review) (average 2.5). This indicator was dedicated to an external review of the model process, which was not done so far. The results were provided by the same group who created the model, with conflicting evaluations (one "moderately dissatisfied", one "not apply", and one "moderately satisfied") and a comment that does not lead us forward (comment 62). This indicator can make sense in a later stage or in another study.

Criterion 25 (Walkthroughs) (average 2.5) This indicator was dedicated to an external group review of the documentation, which has not been done so far. The results were provided by the same group who created the model, with conflicting evaluations (one "moderately dissatisfied", one "not apply", and one "moderately satisfied"), and a comment that reinforces this step was not done (comment 63). This indicator can make sense in a later stage of development or in another study.

Criterion 26 (System Improvement) (average 2). This indicator tried to measure if the system improved by using some outputs of the present study. The results are two "moderately dissatisfied", one "not apply", with two comments, one that the system is not ready to provide results (comment 64), and the other questioning how to use this indicator (comment 65). It is difficult to see systems change at this point of maturity. This indicator was present just in case the modelling exercise brought some very interesting knowledge that could start some small changes, even if in the research agenda, of the system during the study time. It is expected that a positive answer in this indicator should be extremely rare.

Overall performance in the administrative, review, and overview dimension was positive with many "not apply", revealing some steps in the modelling process can be better evaluated in later stages of development. Some topics of interest are:

- Updating the model is time-consuming.
- The efficiency of the modelling exercise could be improved.
- The modelling process is well-documented.
- The capacity for reproduction of the model was questioned due to the need to access information that is not documented, the absence of a formal procedure to conduct a network analysis, to identify the most relevant elements and loops, and for some "overriding command" necessarily taken for coherence.

8.5 Conclusions of the Validation Topic

The process developed here had many positive results, e.g., the list of topics of interest to the developer team. In addition, some spillover effects on the user team rationality were seen, as they documented much of their experience (in concluding remarks) due to the evaluation character of the modelling exercise. The present protocol was applied after months of discussions and many meetings (see section 6) and emails, showing the evaluation process was rich and captured much of the feedback from the user group.

The four dimensions of the evaluation process (Guidelines and Processes, Specific Model Tests, Policy Insights and Spillovers, and Administrative, Review, and Overview) presented different mosaics of positive, neutral, "not applied" and negative evaluations that revealed a rich picture of the process. Some of the most positive aspects, such as the stakeholder interaction with researchers, the documentation of the modelling process, its purpose, and the meaning of the whole process reinforce the importance of the simple SES as a relevant and meaningful tool for the integrated analysis of coupled human-nature systems in coastal areas. In addition, the negative evaluations and comments provided the substance where some adaptations might be required to reach the final goals of the modelling process. Finally, these comments and evaluations provided many topics of interest that can be considered when tailoring the framework to the user needs. A list of these topics was presented after every dimension of the validation process and will be reproduced in section 9 (Conclusions) of this document.

Lastly, some of the indicators of the validation protocol showed to be premature at the present stage of development and might prove to be useful in a later stage, as expected when the creation of the presumed utility protocol.

9 Conclusions and recommendations from the first application of the sSES

We have organized this section by summarizing (i) the main lessons and findings from each DA, in applying the sSES approach; (ii) the validation process, with some lessons learnt to improve the approach; and (iii) some commonalities and recommendations that emerge from the analysis, which can be useful for the future, and especially for improving the sSES approach and potential scaling-up of the process and approach.

9.1 Main lesson and finding from each DA in applying the sSES approach

In the case of the **Tuscan Archipelago DA**, the dual role of tourism in driving economic growth while potentially stressing marine ecosystems and local societies is emphasized. The aim of this DA is to find sustainable solutions balancing tourism with conservation, especially for seagrass beds. We have identified the impacts of tourism on island inhabitant welfare, ecosystem services, marine processes, and human needs, considering the archipelago as a single system despite its island diversity. However, data challenges were significant due to the lack of long-term, high-resolution datasets tailored to the archipelago. This is why indirect indicators were used to compensate for data gaps.

A CLD, representing 26 elements and 53 links, was developed, identifying feedback mechanisms and illustrating complex relationships within the sSES approach. Stakeholder validation highlighted strengths in understanding but noted challenges in replicability and interpreting loop implications.

The key findings and conclusions for the TANP have been:

- Data scarcity poses a significant challenge for SES modelling, highlighting the need for targeted datasets tailored to the archipelago's unique conditions.
- The manual process of updating the CLD with new elements or connections slows down analysis, hindering its efficiency, and suggesting a need for automation tools.
- Stakeholder validation showed high satisfaction with the purpose and engagement of the CLD process but noted challenges in replicability and interpreting loop dynamics.
- Improved communication of CLD results and the ability to infer system behavior under extreme conditions require improvement, suggesting a need for refined guidance in CLD analysis to enhance the applicability and effectiveness of SES modelling in the archipelago.
- In summary, the study underscores the importance of refining data collection methods, automating analysis processes, increasing the speed of the process, and enhancing communication strategies to improve SES modelling and management strategies in the Tuscan Archipelago.

In the **Macaronesia DA**, while the original plan included marine conservation aspects such as MPAs, marine wetlands, and ecological corridors, the study of sSES narrowed its focus to the interactive effects between tourism and MPAs. Hence, using the DAPSI(W)R(M) framework, the study carefully selected indicators across various levels of analysis, considering the economic and ecological values of MPAs and the potential impact of human activities, such as the tourism pressure, on these areas. Data availability posed challenges due to spatial disparities and diverse regional legal frameworks across the Macaronesian archipelagos, with efforts to standardise information and select representative data sources. Establishing causal relationships between indicators revealed discrepancies between quantitative data trends and expert judgment. In cases where trends contradicted expectations, expert knowledge was used to override quantitative metrics, highlighting the importance of expert input in model creation.

The CLD for Macaronesia consisted of 18 elements and 51 links, with ecosystem services and marine processes identified as central elements. Unlike the Tuscan DA, where tourism was primarily viewed as a driver, Macaronesia perceived tourism both as a pressure on marine ecosystems and as a source of economic and societal benefits. Analysis of CLD links highlighted ecosystem services and marine processes as the main elements with significant influence on the system. While ecosystem services are not directly manageable, activities such as recreational diving can indirectly influence marine processes, suggesting the importance of managing human activities for ecosystem conservation.

The key findings and conclusions for Macaronesia are:

- Data scarcity poses challenges in representing the diverse Macaronesian archipelagos, emphasizing the need for careful data selection and interpretation.
- Expert judgment plays a crucial role in reconciling discrepancies between quantitative data trends and expectations, especially in establishing causal relationships.
- Ecosystem services and marine processes emerge as central elements in the Macaronesian marine conservation system, indicating their critical role in ecosystem management.
- Tourism is perceived both as a pressure on marine ecosystems and as a source of economic and societal benefits, highlighting the complexity of its impact on marine conservation efforts.

In conclusion, the study underscores the importance of expert input, careful data interpretation, and the central role of ecosystem services and marine processes in guiding marine conservation efforts in Macaronesia.

In the **Arctic DA,** the ISA was applied to analyze commercial pelagic fisheries. It emphasizes the selection of indicators, timeframes, and the process of data analysis. Hence, indicators were carefully chosen to represent various aspects of the Arctic marine environment and its subsystems, such as the Faroe Islands, Iceland, and Eastern Greenland, with timeframes varying between regions, which reflects the specific socio-economic contexts and historical developments. However, some challenges in indicator analysis have been identified: (i) selecting indicators involved considering data availability, relevance, and the ability to represent complex ecological and societal interactions; and (ii) difficulties arose in determining relationships between indicators due to fluctuating data and the qualitative nature of certain variables.

Regarding the sensitivity matrices built, completing them to assess relationships between indicators proved challenging, with subjective interpretations and potential misleading conclusions for the different subsystems. The process highlighted discrepancies between theoretical expectations and observed trends, raising questions about the effectiveness of the methodology. Hence, the analysis revealed mismatches between expected and observed relationships, particularly concerning employment rates, catch volumes, and ecosystem services. The complexity of ecological processes and socio-economic dynamics led to uncertainties in establishing causal links between the selected indicators.

The CLD analysis, using the Kumu application in the context of sSES research in the Arctic region, identified feedback loops within the SES to understand the complex interactions between different indicators. As such, the relationships between various indicators such as population growth, economic contribution, CO_2 emissions, and fish catch were identified for each subsystem.

The key findings and conclusions for the Arctic are:

- There are some complexities of applying an ISA to analyze commercial pelagic fisheries in the Arctic, since it underscores the challenges in selecting indicators, interpreting data, and identifying meaningful relationships within the ecosystem. The discrepancies between theoretical expectations and empirical observations highlight the need for further refinement in methodology and a nuanced understanding of Arctic marine systems.
- The Kumu software initially crashed due to the computational demands when trying to detect loops. To address this, relationships between indicators that lacked theoretical support were removed, reducing the number of connections and allowing for automatic loop detection.
- There are some concerns about the logic of the identified loops, as well as about mislabeling and misleading interpretations, having some challenges in accurately representing the complexity of sSES and the limitations of data and expertise.
- To improve the process, it should define clearer criteria for identifying indicators, conducting time-series analysis for assessing indicator correlations and ensuring expertise in sSES research.
- It is important to minimize uncertainty and ensure expertise in data analysis before making policy recommendations based on loop analysis results.

9.2 Main conclusions from the validation process

The validation of the sSES approach yielded insights and feedback from modelers, aiding in its refinement and improvement. The evaluation criteria included user satisfaction and commentaries, targeting enhancements for future iterations. The evaluation encompassed four dimensions: (i) Guidelines and Process, in which mixed evaluations revealed areas for improvement in model clarity, usefulness, and presentation; (ii) Specific Model Tests, highlighted challenges in structure verification, loop polarity, and boundary adequacy, suggesting the need for clearer guidelines; (iii) Policy Insights and Spillovers, identified potential policy insights but emphasized the need for further development and community acceptance;

and (iv) Administrative, Review, and Overview, acknowledged positive documentation but raised concerns about model reproducibility and system improvement. Hence, it is concluded that:

- The need for enhanced clarity and usability are crucial for effective model communication and stakeholder engagement, including the relevance of an initial group deliberation as an introductory step. Also, the stakeholders showed a great affinity with the modelers research group but stakeholders experience in the process was limited due to their absence during the model phase.
- Clear guidelines for model structure and data integration are essential for accurate representation and policy relevance. Hence, the question of the adequacy of the model to be used, communicated, or understood by stakeholders should be considered.
- Regarding the approach itself, the tables, the data, the language, and the CLD are complicated to interpret. Also, the limitations of connections (probably the link-polarity or link intensity attribution) should be considered. The selection of variables was made mostly focusing on those that have an associated metric, as opposed to those considered relevant for the system.
- Users excluded relevant indicators due to: a) lack of data, b) problems on how to connect them to the other elements of the CLD, c) incongruence with expert knowledge, d) different time/spatial frames e) lack of computational capacity, f) to avoid data gaps g) incapacity for connecting indicators at the same level in the sSES approach. Some users are connecting variables based on their BOT, and not based on causality. In general, loop analysis is very time-consuming; there are too many loops to consider; users reduced their number of variables to fit Kumu´s capacity. As such, the users need guidance on how to establish geographical boundaries and on the integration of different data sets, governance systems, goals, etc. Guidance on how to create an "overall representative DA" is necessary for consistency. The model is moderately specific" and probably difficult to be applied in other situations with minor changes.
- Further development is needed to translate insights into actionable policies and to gain community acceptance. As such, targets for policy development were potentially found. However, some findings should be considered, since some users commented that results rarely point to recommendations that are not previously known, and the capacity for reproduction of the model was questioned due to the need to access information that is not documented, the absence of a formal procedure to conduct a network analysis, to identify the most relevant elements and loops, and for some "overriding command" necessarily taken for coherence. Therefore, users indicate that results can be potentially transferable with new efforts in modelling and data acquisition.
- Continuous refinement and documentation are vital for model reproducibility and system improvement, although updating the model is time-consuming and the efficiency of the modelling exercise could be improved.

9.3 Some commonalities and problems identified

After considering the works done in the three DAs, to test the sSES, as well as the validation process, **some commonalities and problems emerge**:

- There is **needed a broader knowledge** about systems science **at the first stages of the modeling process**. Also, the user needs to understand the assumptions and flaws, as well as specific differences from traditional statistical modelling to provide an understanding of correlation/causation differences. It is necessary to guide the user on how to integrate and deal with plural instances of the SES such as incompatible datasets, different governance regimes, multi-scale governance and others.

- The approach can be applied, but **data scarcity or quality of data** (including clear criteria in the selection of indicators and timeframes) **should be considered** when applying it. Hence, careful data selection and interpretation are necessary. Guidance on how to integrate, connect, and state the strength of connection of relevant information independently of the data is crucial.
- Regarding the **CLD analysis, the manual process of updating it with new elements or connections slows down the analysis**, and some automation tools should be considered. Here, the expert judgment is crucial, especially in reconciling discrepancies between quantitative empirical data trends and theoretical expectations, and in establishing causal relationships. Overall, **conducting CLD analysis has some complexities and presents challenges in sSES approach research and emphasizes the need for careful consideration of data, expertise, and methodology**.
- Feedback from the **DA applications found barriers with the process related to communicability and replicability** across different contexts for inferring system behavior under extreme conditions without a formal analytical procedure.
- **Opportunities to improve the transferability to managers and end users of the approach were presented within this initial application of the sSES**; these complexities for end users were illustrated by the example of MPA managers in the TANP. This highlights a delicate balance between the CLD comprehensive nature, which could be overwhelming, and the risk of oversimplification when focusing on individual components. **The communication adequacy of the model was also questioned** due to its high complexity and the difficulty in interpreting the tables, data, language and the CLDs by the scientific community and non-scientists alike.
- From the results presented from each DA, there is a possibility that each group provided a varying quantity of work to the tasks involved in this D4.1. With ideas of fair participation in future applications**, it is important to guide the users in the expected amount of work from each area**, to minimize discrepancies and to enable the task to be realistically integrated into work plans.
- The structure of the **resulting model showed many limitations** as the presence/absence of connections, their strength, presence/absence of indicators, exclusion/inclusion criteria for indicators, and use of correlation/causality to establish the connections that summed up to a limited trust in the results.
- **A positive experience with stakeholders was documented**, but from earlier steps in the process, rather than during the modeling phase where they were not present.
- Regarding the use of the results, comments showed **the loop-analysis and selection was very time-consuming, with excessive loops to be analyzed**, a lack of positive criteria to select which of the loops were relevant, the absence of novel insights from the results, and the necessity of improving the efficiency of the process. In addition, **users considered the overall experience positive for learning and some had possibly identified targets for future policy recommendations**.
- The method implementation needs further guidance on how to reduce the input of a large number of loops, to avoid distracting the user with not relevant loops.
- When considering the replicability of the process, users considered it well documented, and **potentially can be applied in other sites, with considerable efforts**. Some “override steps” to establish coherence with specialist knowledge were reported, calling attention to fragilities in the model structure creation.

9.4 Recommendations for improvement of the sSES approach

Considering the problems identified, some points can be extracted from the meetings with stakeholders and partners, their feedback from the validation protocol and from the conclusions that might suggest a way ahead to improve the sSES towards producing a more robust and trusted output and outcome. In addition, they suggest ways to improve the guidance, especially given that eventually the sSES method has to be stand-alone and independent of hands-on support from the developers in order to be rolled-out to other areas. The list below, therefore, represents brief recommendations that can be followed by the developers at their own discretion.

- **Start the exercise expending time discussing the DA users' interpretation of what the problems are, what the physical definitions of their system are, what knowledge the current "system thinking approach" might provide**, that they are as yet not aware of. Discussions with users of their expectations from the tool and the process are necessary to address its limitations. Further explanation and training from WP3 are required for users with natural sciences background using this system. In addition, limitations in terms of governance, participation, political views, governance, etc., must be discussed and its relevance for the result of the model must be explicitly understood. This step contributes to solving problems related to:

    - Integrating multi-level information,
    - Integrating information from different regions, governance systems, timeframes, etc.,
    - Model clarity, usefulness, and presentation,
    - Usefulness of the whole approach.

- **A clear knowledge transfer method should be used to bring the users to the system-thinking modelling body of knowledge. They need to understand why causation and correlation are not the same, and why this method uses the former**. It is also crucial to understand why CLDs are qualitative models and the pros/cons of this method. A clear understanding of how to connect elements should be provided. This step contributes to solving problems related to:

    - Model clarity, usefulness, and presentation,
    - Lack of meaning from the CLD,
    - Lack of trust from DAs,
    - Structure verification, loop polarity, and boundary adequacy.

- **There is further training requirement to show that there is a difference between CLD and sSES per se, and that greater advice is needed in interpreting the former and incorporating these results in the sSES.**

- **As deriving causation from correlation is known to be problematical, the behaviour over time graphs should not be used to connect variables. In addition, the users should not ignore relevant variables due to the lack of data.** Much guidance is available on how to create a CLD (e.g., Sterman, 2000). This step contributes to solving problems related to:

    - Data Scarcity,

- The model being conflicting to user´s experience,
- High number of loops,
- Lack of meaning from the CLD,
- Lack of trust from DAs,
- Structure verification, loop polarity, and boundary adequacy.

9.5 Way forward – towards a new iteration

The results of the testing of the sSES described in this deliverable, and the many recommendations, including those showing dissatisfaction with some aspects of the method, are sufficient information to justify an extension of the baseline assessment process that was initially indicated in the project proposal and timeline as part of Task 4.1. This extended process includes a new modelling iteration and hence enable the realization of upcoming project tasks that depend on such baseline assessment. **This approach (described in detail in Section 10), includes building CLD models again, but this time using expert judgment of the DA representatives (as experts).**

10- New iteration in the process: Building new Causal Loop Diagrams with DAs

10.1. Introduction

Considering the conclusions raised in the first iteration of the sSES application, as demonstrated in Section 9, the additional and complementary to the sSES modelling approach, suggested and developed by the WP4 team, served the purpose of producing the baseline assessment of the DAs SES, that would have specific characteristics, required in order to proceed with upcoming project tasks.

The suggested complementary modelling approach is based entirely on DA expert knowledge and judgement for the creation of CLD models reflecting each DA's SES. It is applied in a workshop format and is divided in four broad sections: and following the agreement made in Zandvoort, a set of workshops and questionnaires were made with the scientists from each DA. The description of the approach, methods, and results are the core part of this section of the report and aim to fulfill the requirements of the **baseline assessment model of D4.1, as a new iteration of the process**. In the present section, four categories of results are presented, namely: a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol. **These four types of results complement and integrate each other and are integral parts of the baseline assessment of the Social-Ecological models as required by the DoA of Marine SABRES.**

<u>**WP4 recommends using the results from the present section to proceed with the work in the other WPs of the Marine SABRES.**</u> This happens because the results in this section are:

- **Meaningful**, as can be seen from the many positive comments in this regard made by the DAs in the follow-up questionnaire (Appendix 2) and the validation protocol (Appendix 3).
- **Robust**, as revealed by the great prevalence of positive aspects and very low rejection of the present model, captured by the validation process (end of this section and Appendix 3).
- **Integral**, the present models embrace the Artic DA and its issues in one model for the whole region. In addition, the Macaronesia DA was modeled as a whole, where both tourism and corridor issues are covered in one model.

10.2 Methods

The new iteration was applied via workshops facilitated by the WP4 team and attended by the DA representatives as experts. Aiming at reducing the risk of an experimental approach, and well-established methods in the systems modeling field were followed. Also due to limited time available to develop proper "guidelines and protocols", the guidelines from Sterman (2000) for the CLDs definition and good practices, and more broadly the guidelines from Van den Belt

(2004) for structuring the modeling process were adopted, from preparation to follow-up. These approaches were described in a methodological plan, that was internally validated in the WP4 lead, AZTI. The exercise includes:

Objectives: to conduct a group model-building process with the three DAs of the Marine SABRES project, focusing on the creation of meaningful causal loop diagrams (CLDs) that represent the issues in each social-ecological system (i.e., tourism and its relations to *Posidonia* in Tuscany; pelagic fisheries in the Arctic; and tourism and the ecological corridor connecting the archipelagos in Macaronesia).
Audience: Marine SABRES scientists' group of experts from each DA (Table 33)
Facilitator: Bruno Oliveira on behalf of the WP4 team
Observers: Angel Borja, Sarai Pouso, Zacharoula Kyriazi and Emma Verling

Table 33: Description of the time and dates of each workshop (number includes observers)

Demonstration Area	Date	Number of participants
Tuscany	2nd and 3rd of May 2024	6 and 4
Macaronesia	7th and 8th of May 2024	11 and 14
The Arctic	29th and 30th of April 2024 16th of May 2024	10, 11 and 10

Expected Outcome of the modeling exercise: A qualitative model represented by a CLD for each DA, which embraces their understanding of the issues related to their DA SES to be dealt with in Marine SABRES and fulfils the required characteristics as mentioned above.

Specificities about the methods: The overall plan for the workshop organization is described by the three-step approach for Mediated Modeling (Van den Belt, 2004). These steps represent broad guidelines that require specific adaptations, as the focus here is to produce a qualitative CLD as methodologically described in Sterman (2000). As this step of the project was not dedicated to produce a manual on the creation of participatory models, a generic procedure is described where the process is divided into three steps (A. Preparation, B. Workshops, and C. Follow-up):

A. Preparation
For preparation, the most relevant literature about system thinking was sent to the DAs to collaborate with the learning process of the system dynamics field (e.g., Meadows, 2008; Sterman, 2000). This previous understanding of systems science was limited due to the short time for the preparation phase. To help with this limitation, an introductory learning section was provided on the first day of the workshop to explain and discuss relevant topics for building CLDs, some general literature on systems thinking, and some papers with examples of CLDs for the supposed problems of each DA (e.g., Crabolu et al., 2023; Gercek et al., 2022; Käll et al., 2022).

The definition which group should be present in the modeling process (the audience) was established by the project previously as being the group of scientists that are representative of each DA. We recommend, and most of the DAs reinforce this recommendation, that a broader group of stakeholders be incorporated into the modeling process in the future as they bring invaluable knowledge and experience from different perspectives, views, and interests about the systems issues, and possible innovative solutions to the problems at hand. This message is strongly reinforced by the comments of the DAs during the validation step (Appendix 3).

As this is not the first time the groups have reunited to produce a model, preliminary discussion focused on the constructive and meaningful way ahead towards producing a CLD that qualitatively represents the problems in each DA. It was not the scope of these workshops to discuss previous experiences in modeling but to concentrate the effort on the production of the CLDs.

B. Workshops

The workshops were structured in three steps (adapted from Van den Belt, 2004):

B.1. Introduction

This step started with a discussion of the programme for the workshops, and a short description of the CLD-type model, showing their limitations and qualities. It then opened to a group discussion of how this qualitative model could contribute to their problems. Also, in this step, the “ground rules” for the workshop were presented and agreed upon.

B.2. Problem articulation

This session was crucial and is described in results in terms of Q&A (when necessary for grammatical coherence, some words were added to the original text and are in brackets). It was conducted by an open discussion on the issues of each DA, specifically considering the limitations of the present approach in terms of geographical boundaries, participation and representativeness boundaries, time horizon considered in the model (past and future); probable main variables that might contribute to explain the behavior of the system. This problem articulation session also discussed which problems they wanted the model to help with and what was the purpose of the exercise. In addition, the implications for policy-making processes and comments about the users of the model were made, considering the constraints of the preset exercise, but to produce a qualitative model, with limited complexity but seen as meaningful by the participants.

B.3. Qualitative model building

The qualitative model-building session was conducted to explore the views of the DA scientists regarding the problems identified in the previous section and translate them into a causal loop diagram. Alternate and iterative steps of the “feedback loops approach” and “modular approach” (Van den Belt, 2004) were conducted freely to explore which elements were perceived relevant to the problem at hand and their connections. The CLD model was constructed using Vensim PLE x64 (version 10.1.0). Participants were encouraged to describe the connections of the system at a DA level, and considering the restrictions explored in the problem articulation session (as time and geographical boundaries). Some variables that were considered irrelevant were excluded concomitantly with the building process by consensus of participants.

C. Follow-up

By the end of the modeling process, the groups were presented with three sets of questions. These were related to a) a validation protocol; b) a pro-environmental/worldviews questionnaire and c) general integration of results with other WPs.

The validation protocol is the same as was described in Section 5 and Appendix 1 of the present report. Results are described in the validation section.

A pro-environmental behavior and worldviews questionnaire was presented to provide some guidance on the intrinsic bias that was being integrated into the models, as it was a group-building process. This exploratory topic can be integrated in the future into D4.2 which will deal with scenarios, as worldviews are part of previous deliverables of Marine SABRES (Bremner et al., 2024; Bremner & Oliveira, 2024).

General questions regarding the integration of the results of the present exercise with other WPs and deliverables in the project were considered an opportunity for the DAs to collaborate with respect to their views on the topics and to understand to which extent previous knowledge from other WPs and tasks was integrated into the present exercise. The answers are transcribed here as a Question and Answers topic, as we thought this format would be the best to represent the original answers made from the DAs to the specific topics dealt within this session. WP4 chose to not "complete" phrases or attribute the name of the speaker to each sentence to keep the originality of the contribution, and to keep these results as reflecting the position of a group, not individuals. Results are described in the follow-up and integration section.

Finally, users were invited to continue the modeling process by envisioning their DAs concerning scenarios. As the previous products in Marine SABRES regarding scenarios were focusing on multiple possible futures (e.g., SSP1 and SSP2) not to be exhaustive, there was a necessity to discuss these plural futures with the DAs. This topic will be explored in depth in the deliverable D4.2.

Limits and Caveats

This work was prepared and executed within a short time horizon. Therefore, obvious limitations are present, such as the lack of an *ad hoc* manual for the cases in Marine SABRES on how to define the scope and produce the CLDs. Other clear limitations are regarding the nature of the model created in this case, a qualitative causal loop diagram, that is useful to promote a systemic understanding of the issues in the DA (as stated in Appendix 2), but with limitations to answer their aspirations for a precise numerical model. WP4 sees the selection of this kind of model as representing a timely and novel step into the system thinking from the consortium and does not imply that future numerical models cannot be derived from the present qualitative model, exploring numerically what is qualitatively showed in this section. In addition, the comments received from the DAs, which are the users and the focus of the modeling exercise, also bring relevant contributions to the improvement of the techniques used here, especially with regards to the consultation of stakeholders, the maturity of the model, etc. These comments will be addressed, in what is feasible and possible, when the content of this section is prepared as a publication.

10.3. Results from the CLD creation for each DA

The results from the modeling exercise are presented in this section. They are organized into **four categories** of results, the product of the organization of the workshops as described in the previous methodological section.

First, the **problem articulation session description**, which shows the description of the problems, limitations of participation, geographical boundaries, periods, purpose, expected users, and policy implications of this modeling exercise. This section builds on the DoA objectives for each DA and explores the vision of the scientists about the specific issues of the DAs. This is relevant to WP5 as it shows limits and opportunities to be explored in the system and provides an understanding of the expected governance-related aspects of using the knowledge in the model.

Second is the **CLD model** and its description, showing the material result of the model exercise. It connects the elements that are relevant to understanding the issues described in the previous section, from the DA scientist's view. All graphs have some variables identified with one of the PESTLE elements before their names (e.g., "(En) Habitat Quality" in Figure 45). These elements represent the Political, Economic, Social, Technological, Legal, and Environmental (PESTLE) variables identified by the DAs as the most relevant parts of the system for management. This result reveals targets for management and interventions in the system and fills the requirement of the **major components of the system function**, requested by the DoA. These variables, their systemic properties, loop influences, archetype behavior, etc., will be explored in-depth in D4.2 to provide the pathways in the DA system toward desired scenarios. The present result is relevant to WP5 as the graph can be used for many types of simulations (T5.3), the descriptions of the current state and future conditions (to be explored in D4.2) will feed in T5.2 as the counterfactual situation and pathway for the "cost analysis" that will be done in the socio-economic scenarios.

The third category is the content from the **follow-up and integration questions** made after the modeling exercise. This part brings the DA understanding of the message the model is passing, the recommendations they can expect at this stage of the modeling exercise, and the extent of the novelty to which they were exposed during the present process. In addition. this section brings seed ideas to be explored as possible alternatives to the management of the regions (or **initial concrete measures** as required by the DoA). Furthermore, to connect the present modeling with previous stages of stakeholder investigation in the project (i.e., D2.1), the DAs were asked about the potential acceptance of the knowledge presented in the model by their stakeholders, the relevance of the knowledge in D2.1 for their views of the system, how much of the knowledge in D2.1 was specifically used in the modeling and how this previous knowledge could be integrated with the product of these workshops.

This part of the present baseline assessment report is useful for integration of the present results with WP2 products, and also relevant for WP5 to see ideas of alternatives for management. In addition, it complements the next session (validation protocol) with statements of the satisfaction the DAs have regarding the modeling exercise. We had more than **100 statements** describing these topics (Appendix 2) that were suppressed from the main text for simplicity, but are invaluable to the audience interested in the DAs view of their areas.

Lastly, the fourth category is the results from the application of the **validation protocol**. This protocol was used as an organized form of collecting information from the group involved in the modeling exercise regarding their level of satisfaction with the process, their views of the limitations and qualities of the process, the boundaries of the model, etc. The protocol is described in Appendix 1 with the literature on which it was based. Here we bring the general results for this step and comments on each criterion. In this validation exercise, we had almost **300 comments from the DAs.** The results from the validation protocol are important to consolidate the CLDs of this section of D4.1 **as robust results from WP4 to be carried out and explored in further analysis and WPs of the Marine SABRES project.**

10.3.1 Macaronesia

Problem articulation session

Q: What are the issues of the DA, specifically considering the limitations of the modeling exercise in terms of geographical boundaries?

A: MPA and Tourism and the corridor connecting the archipelagos. The challenge is to find a common ground between the archipelagos. The geographical delimitation is a challenge because the Azores and Madeira are different autonomous regions [in one country (Portugal)], which with the Canary Islands [(another autonomous region and country, Spain)] form a difficult and heterogeneous region to model. The pressures (and levels [of pressures] for instance, created by Tourism) are different in each area. The distance between these areas is a limitation for the scientists to discuss and find common ground. Biologically [the common delimitation] is interesting between the similarity of species, there is the similarity of being oceanic and volcanic islands with deep-sea areas, similar currents, etc. Politically it is challenging because the DA has two countries and three autonomous regions. Example: Wetlands and seagrass are not evenly distributed. Restoration is a topic that appeared by the project development. Offshore MPA for the Azores can be discussed in terms of biological data provision for future assessments. There is a natural isolation from marine organisms of the Azores in connection with the rest of the DA. The idea of the corridor was based on megafauna more than benthic smaller fauna. There is a cultural heritage shared in the DA-Additional topics: climate change, invasive species, and corridor based on corals.

Q: How do you consider the participation and representativeness boundaries in this exercise?

A: Most of the present are biologists, and therefore other views are necessary. Someone from the economic (business stakeholders too), political departments, and NGOs would be interesting.

Q: What time horizon we should consider (past and future)?

A: Within the archipelagos a different time frame. Around 20 years for the DA [as a whole]. Azores, Madeira intermediate, but the Canary Islands have a more antique tourism activity. There are differences in spatial and time frames. Pre-covid and post-covid are also relevant topics. For the corridor, the idea is to consider between 5 and 8 years in the future. The time frame for the corridor might be extended more than 20 years in the past.

Q: What is the purpose of this exercise?

A: How to integrate different views and knowledge regarding the management of the DA and its problems. To incorporate the social and stakeholder part of the system, to learn how people are receptive to management actions that lead to a sustainable management or future of the region. Identify common problems and promote the resilience of the DA.

Q: Who should use the model?

A: All the users (MPA managers, residents, tourism operators, users of the sea and the land), fishermen, commercial fisheries, government and policy-makers, NGOs, Academia (such as students, scientists, and others), and opensource everyone interested in the topic. Difficulties in language and terminology can interfere with the ability to understand the models.

Q: What are the expected implications for policy-making processes?

A: It would be interesting to promote the discussion between policymakers from the three archipelagos in terms of common topics of interest and common futures. National policymakers would be required to understand the different boundaries between the systems and understand how decisions could be taken in an integrated form (at the DA level). As the corridor crosses international waters, it would be interesting to have higher-level institutions discussing this as well.

Q: Are there other topics relevant to the model?

A: No.

CLD Model for Macaronesia

The model of the Macaronesia region integrates the two countries and the three archipelagos which is novel in Marine SABRES. In addition, it also contemplates the two most pressing issues identified by the DAs for the region: tourism (also appointed by stakeholders as important in D2.1) and the idea of an ecological corridor connecting the three archipelagos (Figure 45).

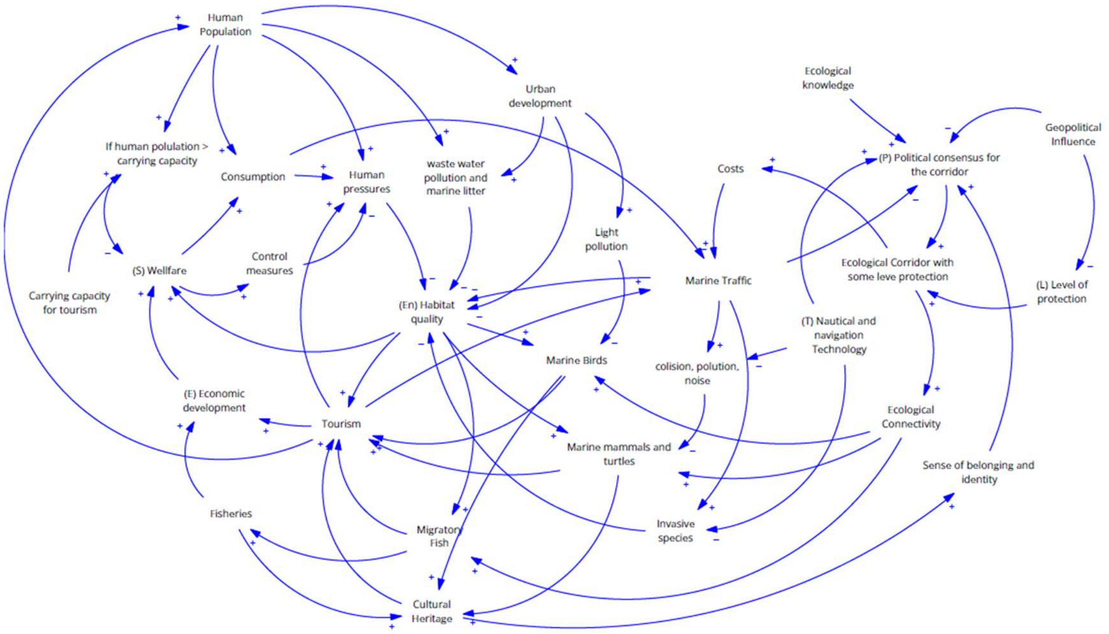

Figure 45:Causal Loop Diagram for the Macaronesia Demonstration Area

The story in the CLD started with human pressures negatively affecting the habitat quality (main environmental variable), which in turn can have many influences, such as on marine birds, mammals and turtles, migratory fish, tourism, and welfare. Marine birds, mammals and turtles are relevant for the tourism activity, but also to the cultural heritage of the archipelagos (with fisheries as well). 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 the goods reach these archipelagos by boat. Then this marine traffic can have environmental influences such as collision with mammals, facilitating invasive species colonization, etc., but also have a negative influence on the consensus for the creation of the corridor (the main political variable), as these vessels would have increasing costs by the creation of the corridor. 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. The presence of the corridor is seen as positively influencing ecological connectivity, which would, in turn, promote the abundance of marine birds, mammals and turtles, and control the presence of invasive species. Nautical and navigation technology (the main technological variable) was seen as facilitating the political consensus for the corridor and reducing the negative impact of collisions, pollution, and invasive species.

Follow-up and integration questions

When asked what main message the model is passing, the responses embraced nine participants of the workshop (9 out of 11). The answers from the Macaronesia DA were extremely positive as they revealed that the group understood the ideas of a system thinking perspective in connecting the elements in the system, its impacts on each other, opportunities for intervening in the system, etc.

The group was also consulted on what policy insights or recommendations could be created meaningfully. That question is known to be a bit premature, as the model is still in the first stages of development, but the answers from the group were positive in revealing some insights to promote sustainable management. This item shows potentially some **initial concrete measures** that need deeper investigation, but as this is requested by DoA, they were copied here:

1. One of the main concerns raised is that an ecological corridor among archipelagos would cross many geopolitical boundaries (e.g. regional, national, international) and that this needs clear **articulation with the different agents**, which may not be easy to accomplish.
2. We did not enter into specific details. We discussed implementation of protection measures in the ecological corridor Azores-Madeira-Canary Islands. From the CLD we can suggest that such measure can be adopted and maybe can be accepted **more likely if technological development is "moving" in the same direction**. For example, the development of a less noisy engine, or detection system to avoid collision with cetaceans.
3. Policy insights, as many as the stakeholders. Not sure about recommendations.
4. Recommendation for healthy ecosystems and habitat quality. Protective measures. Sustainable tourism.
5. Supranational ones concerning a possible marine corridor and its management.
6. **Some insight can be created, such as the rearrangement of marine traffic to prevent environmental impacts, the implementation of innovative technologies for increased sustainability and potential cost reduction, as well as the evaluation of size restrictions and fishing campaigns for fisheries**, based on scientific evidence of size to sex ratio and reproductive activity of some species. Additional insights include **shifts in tourism activities to a more sustainable and local economy-based approach**. Despite the potential of these considerations, further evidence and resources might be needed to back them up.
7. A **recommendation for the establishment of an ecological corridor for Macaronesia, and maybe to propose some transboundary MPAs in the area**.
8. The **increase in regulation regarding over-tourism**, and the **creation of an MPA covering the biological corridor**.
9. The possibility of creating an ecological corridor at a supranational level has the potential to improve policies associated with the management of small island systems and their particularities.
10. The recommendation keywords are regulation, protection, and conservation. Regulate and protect MPAs and the ecological marine corridor considering human pressures, to ensure healthy and sustainable ecosystems

The ideas regarding systems thinking and interdisciplinary methods are new for a great part of the researchers dedicated to marine science. This can be seen in the description of the participation boundaries of the model, where all DAs evidenced the disciplinary limitations of their groups (mostly from biology or related disciplines). Therefore, a question was made to understand how novel was this process to the DAs, which provided positive feedback from their experience with the modeling process.

We asked the DA scientists what their expectations regarding the potential acceptance of the knowledge represented in the model would be by their stakeholders. The answers were positive but highlighted the necessity of considering different stakeholders' perspectives, knowledge, etc. In addition, the DA was asked how they evaluate the relevance of the information of D2.1 in their knowledge base used in the model. The idea was to understand in general, how much of the stakeholder-interview report was used in the present exercise. The answers are positive but reveal a spectrum from positive to ignoring the content. Then a specific question regarding the use of the content from stakeholders' previous investigations was made to understand if the DA used direct information from D2.1 and in this case, most of the respondents said they did not use any direct information from the cited report.

10.3.2 The Arctic DA

Problem articulation session

Q: What are the issues of the DA, specifically considering the limitations of the modeling exercise in terms of geographical boundaries?

A: The fish is what connects [ecologically the DA], the social system (political) that takes decisions on this fishery [is different]. The ecosystem is pretty much the same, but with regionally different cultures. The DA is just part of the system, as lacks the UK, Norway, etc. [the DA questioned] Why did not the project use all countries in the region? [Angel Borja answered:] There was a limited number of partners due to budget restrictions. The definition of the area does not correspond to the ecosystem's spatial limitations. There was an idea of dealing with islands.

Q: How do you consider the participation and representativeness boundaries in this exercise?

A: More people (stakeholders), government, industries, etc. should be included, but it would be difficult (challenge) to have them at these tables. It would be great to have a way (practical and not academic) for people to use it quickly. There is a similar project going on in Norway (UK also?), Maybe in the end we can compare the late stages with other countries. The second limitation is the number of countries present in the discussion.

Q: What time horizon we should consider (past and future)?

A: [We should] go back between 10 to 20 years because of data availability, and distributions of fish changes (stock changes), A Future period of 3 years, if possible, for more time would be interesting. Are we going to be fishing in the next 20 years? Is it going to have fish? (maybe not real for the mackerel, and more realistic herring) (capelin is very unpredictable).

Q: What is the purpose of this exercise?

A: It is the goal of this exercise to have better ecosystem-based management/Governance considerations for climate-driven distribution changes of highly migratory species of the Arctic

Region and other elements (defined as new stocks to be fished, tourism impacts (Reliability for the population to participate in these sectors governance). Questions to guide the discussion: What will the future economy look like? Do the stocks present the stability necessary for the business to jump in? If possible other elements can be investigated such as aquaculture (there are companies interested in this economic sector).

Q: Who should use the model?

A: To present to people we talk to (managers, stakeholders, decision-makers, politicians, fishers, communities, municipalities, companies (social responsibility), ministry). Hopefully not just academically but to understand different ideas between municipalities and companies. The model could increase “legitimacy (including diverse knowledge)/awareness” in decision-making due to their understanding of the system.

Q: What are the expected implications for policy-making processes?

A: [Implications would be to] clarify the whole EBM (be clear that the ecosystem is there!) approach and make it useful to people. Redefine EBM along the exercise. Tipping points can be informative to guide management. Explore the ideas of tipping points/resilience / Panarchy (where are the weak links?). How can the governance of the system support (hinder sustainability) the system? Consider participatory management (as companies, everyone has a participation in the system).

Q: Are there other topics relative to the model?

A: Connections with energy transitions, marine mammals and predators, unemployment (could become a problem/opportunity?) is it related to immigrants? Is it a problem? Are competing activities influencing jobs and the environment (mining)? DSM and offshore oil and gas might be influencing.

CLD Model for the Arctic

The Arctic DA produced a single CLD that integrated the pelagic fisheries issues in the portion of the Arctic Ocean that is understood as a common area for the DA (Figure 46).

The story in the model starts with the agreement on quota allocation (the main political variable). 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. It is also influenced by the variability of fish distribution, and the predictability of productivity, environmental variables that are filled with uncertainties and influenced by climate variability (main environmental variable). As a rational agreement on quota allocation is provided, it influences positively the stock of fish due to the rational and efficient exploration of the resource, reducing the fishing effort, which is related to the landings (fish on land). Landings (the main economic variable) can influence the stock of fish in the long term if overfishing is the case but is important to the exports and local development (which are both the most important social variables). Landings also influence jobs, which summing with profit, are the main drivers of taxes and then 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 (the main technological variable) and then influence back on jobs.

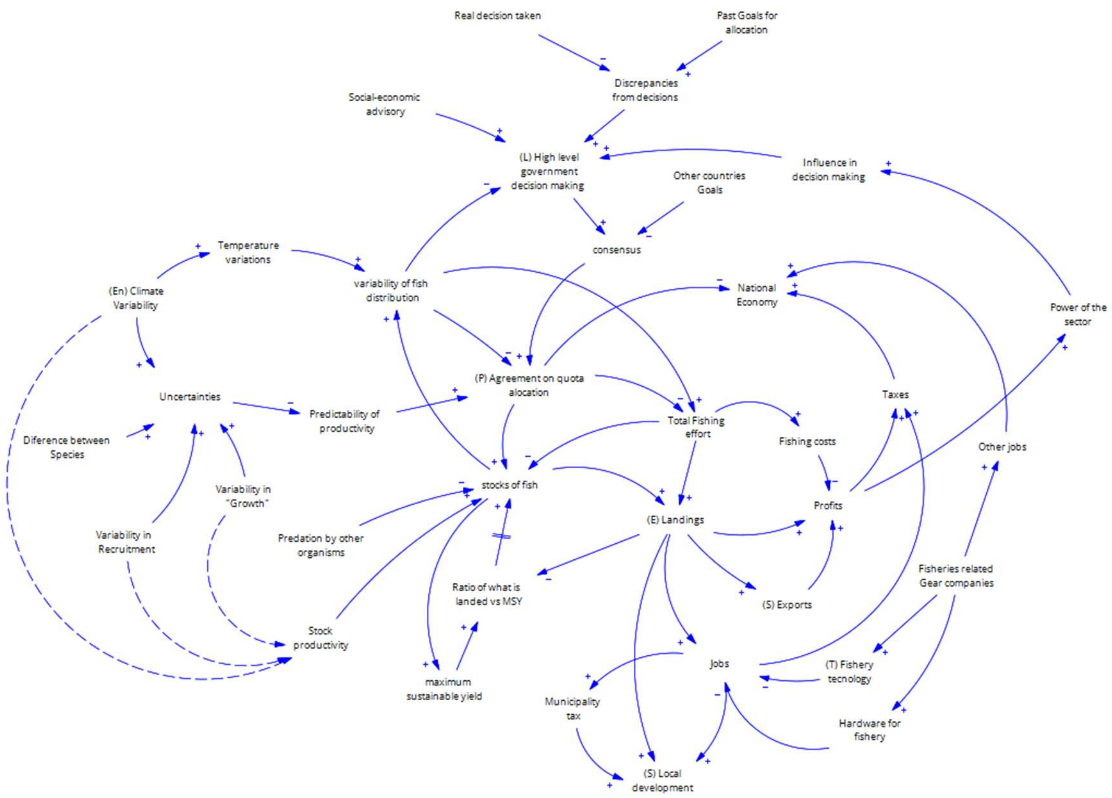

Figure 46:Causal Loop Diagram for the Artic Demonstration Area.

Follow-up and integration questions

When asked what main message the model was passing, the few answers from the Arctic (2 out of 9) were a mixture of one positive understanding of the message, and the other being confused about the naming of loops and the message of the model. The group was also consulted on what policy insights or recommendations could be created meaningfully. That question is known to be a bit premature, as the model is still in the first stages of development, and the answers for the Arctic confirmed that expectation. This DA could not identify any **initial concrete measures** to promote sustainable management at this point of the modeling exercise.

Answering the question of understanding how novel was this process to the DAs, the answers were a mixture of learning about the relations of the impact of fishing on the socioeconomic system with a second view that no novelty was discussed due to time constraints. We asked the DA scientists what their expectations regarding the potential acceptance of the knowledge represented in the model would be by their stakeholders. The answers did not show any potential acceptability of the results to the stakeholders. In addition, the DA was asked how they evaluate the relevance of the information of D2.1 in their knowledge base used in the model, and they answered that the knowledge in D2.1 was not used to build the present model. When then specific question regarding the use of the content from stakeholders' previous investigations (i.e., D2.1) was made, the answers were also negative.

10.3.3 Tuscany Archipelago DA

Problem articulation session

Q: What are the issues of the DA, specifically considering the limitations of the modeling exercise in terms of geographical boundaries?

A: Tourism-related issues, the main idea was to understand the relationship between tourism and the conservation of the coastal ecosystem, particularly the coastal areas' seagrasses (*Posidonia oceanica*), towards obtaining sustainable tourism and the well-being of local communities in the areas. In addition, anchoring effects, seagrasses, increasing boats, and the increase of water nutrients cause eutrophication. The geographical limitation is satisfactorily representative as it includes all the islands, as they represent different realities with a range of different conditions, gradients of pressures and conservation (inside and outside the MPA), and gradients of different management approaches. The mainland is excluded (maybe Livorno and Grosseto could be included) for simplicity.

Q: How do you consider the participation and representativeness boundaries in this exercise?

A: The group is formed by ecologists as background and therefore the view about the system is stronger in the ecological part, with limitations on other parts of the system where people have less specialty. The group is aware that people with different backgrounds would bring a wider view of the system, for example, the divers or fishers can have a different view. Complementary, to previous work with the stakeholders, some consensus was achieved on what are the relevant topics and issues of the region.

Q: What time horizon we should consider (past and future)?

A: The time-scale changes in the system could be perceived in the last 20 years, and the tourism impacts seem stronger from the last decade or less. Climate change topics would need a larger timescale. The national park is from the late 90´s and might be a relevant time window to consider. For the future ideas, 10 years seems enough.

Q: What is the purpose of this exercise?

A: The aim is to understand questions that might be tested in the future, through experiments or other techniques and use it to feed marine managers towards decision-makers. To understand the connections of ecology with other topics (SES) and achieve a better position to advise decision-makers in a good direction. Also, a way to highlight the work done so far and communicate with stakeholders/ other people not involved with conservation. Considering the DSS (that should be quantitative), provide science-based answers to specific managers' questions. Understand the needs in data to transform this knowledge into numerical models. A visualization of the potential of models (and related data) to inform decision-makers and inspire them to use and learn towards "informed" decision-making.

Q: Who should use the model?

A: The model could be ideally accessible to both low-level stakeholders, but also marine managers, ministers, and high-level involved personnel. It could be useful for people who have activities that support tourism activities to be aware of the importance of conservation of the archipelago. The tool might be too complex to communicate and maybe a simplified message (to a broad audience) would be more appropriate, and therefore this must focus more on high-level decision-makers. It could also be useful for the scientist group to use, and improve where they feel necessary, and guide the scientific team on focusing on relevant parts to improve the science development.

Q: What are the expected implications for policy-making processes?

A: Hopefully, we can provide more informed and supported decisions, but there might be limited acceptance by the decision-makers [as they might] need time to recognize the value of this approach, they may have other priorities (conflicting), etc. When the current practices (elsewhere) become the use of DSS, probably Italy will adhere to the common practice.

Q: Are there other topics relative to the model?

A: Restoration can be included in the discussion, but there were limited discussions previously in the project. The 1,000 m^2 of *Posidonia* restoration was applied on the field and will be monitored this year.

CLD Model for Tuscany

The Tuscany Archipelago DA produced one model where the issues relevant to tourism management were integrated with the pressures on *Posidonia* seagrass meadows (Figure 47).

The story in the model starts with the tourism activity. Tourism is influenced negatively by the limitation of tourism (in MPAs) (the main legal variable). Tourism is central in the model, as it influences plastic pollution, boat traffic, and nutrient discharge, but also the local economy, diving centers' income, boat renters' income, and others. Boat traffic leads to anchoring in the coastal zones, which affects the seagrass meadows (the main environmental variable). The seagrass meadows influence carbon sequestration (leading, with delays, to climate change and extreme events), erosion control, and public services, but also fauna and flora biodiversity, health, control 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 overpopulation of tourists. Jobs influence taxes (the main economic variable) which are one of the responsible for public services. An app organizing the tourism activity (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. These nutrient discharges are influenced by tourism activity and boat traffic, are regulated by land-distance norms for discharging, and influence the water quality which inspired the blue-flag program for safe tourism. Nutrient discharge and anchoring affect negatively the restoration efforts that feed the seagrass meadow population.

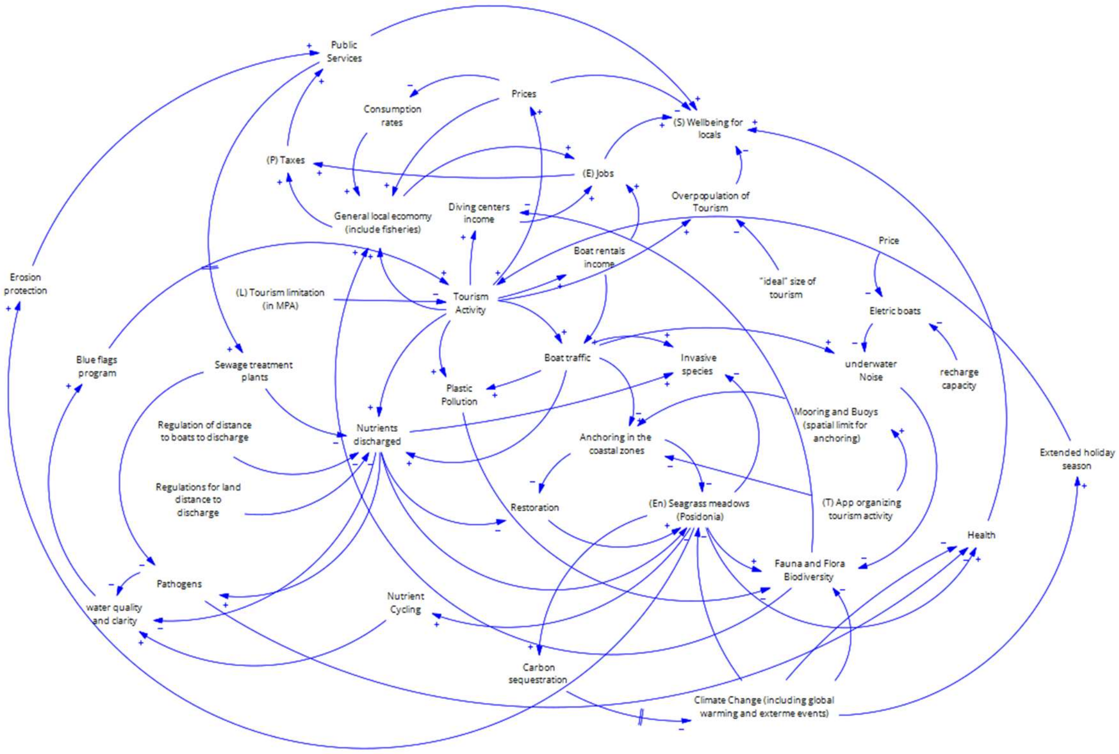

Figure 47:Causal Loop Diagram for the Tuscany Archipelago Demonstration Area.

Follow-up and integration questions

When asked what main message the model is passing, the responses embrace almost all participants of the workshop (3 of 4). The answers from the Tuscany DA were extremely positive as they reveal the model can show the positive and negative aspects regarding the tourism activity, and its systemic connections (feedback).

When the group was consulted on what policy insights or recommendations could be created meaningfully, the group listed potential interventions that could fulfill the **initial concrete measures** requested by DoA. Their answers were:

1. **Regulation of touristic activities within the MPA but also outside**. **Deployment of buoys** in the most visited places/diving sites for private boats, to prevent *Posidonia* damage. Dissemination and educational activities to **inform the public audience about best practices at sea**, marine environment, and species in the Tuscan Archipelago. **Public funding to help locals shift from extractive activities (fishing) to more sustainable tourism-related jobs**.

2. One recommendation that arose from the modeling process was **creating a series of apps to engage with the public** to let them make informed decisions about their tourism activities in the archipelago. Furthermore, **instituting policies aimed at regulating the number of tourists was mentioned**.
3. **Local people and tourists should be adequately informed** about the importance of conserving marine biodiversity and key habitat-forming species. More information about this topic is needed.

As before, the ideas regarding systems thinking and interdisciplinary methods are new for a great part of the researchers, and when questioned on how novel was this process to the DAs, the answers were extremely positive, showing the DA could integrate knowledge and perspectives they did not have before.

We asked the DA scientists what their expectations regarding the potential acceptance of the knowledge represented in the model would be by their stakeholders. The answers showed the knowledge in the model is potentially acceptable, but called attention to considering different stakeholders' perspectives, knowledge, etc. When asked how they evaluate the relevance of the information of D2.1 in their knowledge base used in the model, the answers revealed the knowledge in D2.1 was used as a background for the knowledge in the present model. When the specific question regarding the use of the content from stakeholders' previous investigations was made, the answers revealed that most respondents used the knowledge from D2.1 to the present model.

10.4 Validation for the CLD workshops

These results correspond to the methods briefly described in Section 5 and Appendix 1. The aim of applying the validation protocol in the present CLD production process was to understand what the limitations and qualities of this modeling are. There are limitations of the present approach due to the short period for preparing the workshops, interacting with the audience, previous contact with the literature, etc. (see limitations and caveats section). Therefore, the validation process in the present case is firstly dedicated to understanding if even with the time constraints this exercise was imposed, the models are robust and meaningful to proceed with the project. It is not a goal of the present validation to provide material for refinement, as the present CLDs will not be refined or reproduced. Also relevant to highlight, we had an overwhelming number of evaluations (more than 400 numerical evaluations and almost 300 comments in this section plus 100 comments on the "follow-up and integration" section), and therefore these results will be commented on in aggregate. There will be a shorter version of the validation exercise showing all the comments that will be turned into a paper. All comments are available for consultation (Appendix 3).

The overview evaluation of the process of modeling (Figure 48) was very positive with a very low number of evaluations being of "very dissatisfied" and "moderately dissatisfied", which reveals a very low (7%) dissatisfaction with the process against "moderately satisfied" (27%) and "very satisfied" (10%), revealing the overall satisfaction (37%) with the process. Answers that considered the item "do not apply" were (32%), and 24% were neutral ("neither satisfied nor dissatisfied") (23%), summing the 416 evaluations. These individual evaluations have variations, and therefore the consistency of each quantitative statement must take this variation and the specific comments into consideration.

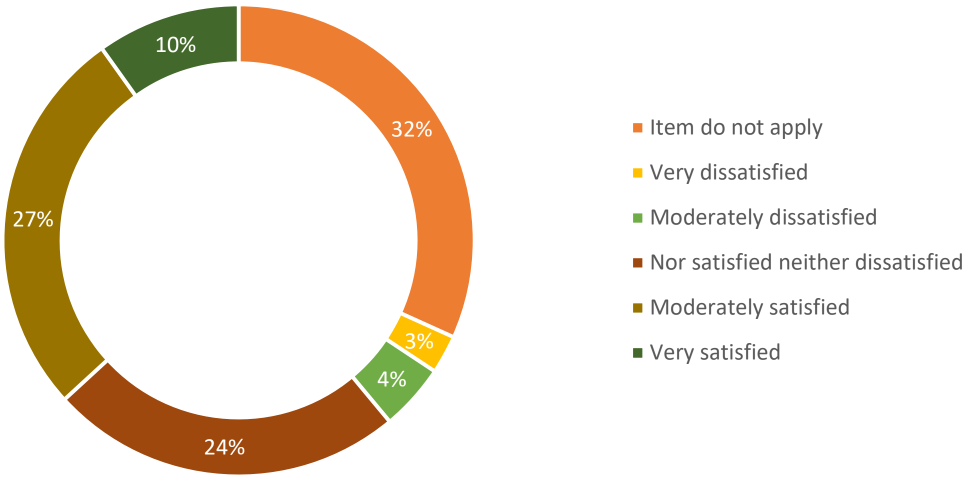

Figure 48: Distribution of evaluations in the validation of the Causal Loop Diagrams.

The four types of indicators ("Guidelines and Processes", "Specific Model Tests", "Policy Insights and Spillovers", and "Administrative, Review, and Overview") present different levels of satisfaction and dissatisfaction that point to different aspects and steps of the modeling process. This division into four dimensions of the validation protocol allows a nuanced analysis of the process from the start (guidance and purpose); of the represented indicators, model assumptions, boundaries, variables, etc. (specific model test); an evaluation of the capacity of the model to create some spillover or change in the system (policy insights and spillovers), and, to its documentation and replicability (administrative, review, and overview).

The distribution of results from the evaluation of each dimension (Figure 49) is considered very positive by the DAs evaluations due to the negligible dissatisfaction rates, especially in the first two dimensions, the greater number of positive evaluations, and the higher presence of "not apply" later in the third and fourth dimensions.

The positive results obtained in both initial dimensions are considerable, specifically because considering the initial maturity level of the model developed in these workshops, these dimensions represent the most important parts, such as structure, meaning, participation, etc. We encourage reading the comments (Appendix 3) to understand how the DAs experienced the exercise.

Table 34:Tests for presumed utility in qualitative models. A, B, and C are the Demonstration Areas. R1-10 are respondents per DA. The scale is from 0 - item does not apply (white), 1 – very dissatisfied (red), 2 – moderately dissatisfied (yellow), 3 – nor satisfied neither dissatisfied (grey), 4 - moderately satisfied (blue), and 5 – very satisfied (green). The colors are illustrative of the values. Avg is the average of those results which excludes zero. Mo is the mode, when possible. PIoR: Policy Insights and Recommendations.

	N	Criteria	A										B			C				
			r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r1	r2	r3	r1	r2	r3	Avg.	Mo.
Guidelines and processes	1	Purpose	5	5	4	5	3	4	4	4	4	5	2	3	2	5	4	4	3.9	4
	2	Usefulness	4	4	4	4	4	4	3	4	5	4	3	3	2	2	4	4	3.6	4
	3	Presentation	5	3	4	3	4	4	3	3	4	3	3	4	3	0	3	4	3.5	3
	4	Perspectives in Boundary-adequacy	1	3	3	3	3	5	3	3	4	3	3	4	2	4	4	4	3.3	3
	5	Norms/values in boundary adequacy	4	0	3	0	4	5	4	4	4	0	3	4	2	5	4	3	3.8	4
	6	Trustworthiness or Guru status of the system dynamicist	0	0	3	4	0	0	0	3	5	0	4	0	0	5	4	4	4.0	0
	7	Meaningfulness of the process	4	0	4	0	0	0	0	3	5	0	0	0	0	5	3	5	4.1	0
Specific model tests	8	Structure-verification	4	4	4	4	4	4	3	4	4	4	5	3	2	4	3	4	3.8	4
	9	Loop Polarity	5	0	0	0	0	3	0	3	0	0	4	3	0	3	4	4	3.6	0
	10	Boundary adequacy (as structure)	4	3	4	3	5	5	4	4	5	4	5	4	3	4	4	4	4.1	4
	11	Family-member	4	4	4	4	5	4	4	4	5	5	3	2	4	4	3	3	3.9	4
	12	Extreme-conditions	0	4	3	3	0	3	0	4	2	3	4	0	0	1	0	0	3.0	0
Policy insights and spillovers	13	Insight generation capacity	3	3	0	3	3	5	3	0	0	3	1	3	1	3	3	4	2.9	3
	14	Relevance and Fertility of PIoR	4	0	3	0	0	4	0	0	0	0	3	3	0	2	3	3	3.1	0
	15	Congruence of PIoR with culture	0	0	0	0	4	3	0	0	5	0	0	3	0	3	3	4	3.6	0
	16	Boundary adequacy (as policy)	3	0	3	0	3	3	0	0	4	0	0	3	0	3	3	2	3.0	3
	17	Learning	2	0	0	0	4	5	0	0	4	0	3	4	0	4	4	4	3.8	0
	18	Engagement	0	0	0	0	5	0	0	0	3	0	3	0	0	0	3	0	3.5	0
Administrative, review, and overview	19	Ease of Enrichment	4	0	3	1	3	5	4	5	3	3	4	4	1	4	4	3	3.4	4
	20	Time & Cost of the Intervention	5	5	4	5	3	3	4	3	4	4	4	4	2	3	4	5	3.9	4
	21	Documentation	3	2	3	2	5	3	5	3	5	5	3	4	1	4	3	4	3.4	3
	22	Replicability	3	1	2	2	4	3	3	0	4	3	3	3	1	2	3	2	2.6	3
	23	Audit or cross-validation	0	0	4	1	4	5	0	0	3	3	0	0	3	4	3	4	3.4	0
	24	Higher-level Model review	0	0	0	1	5	5	4	0	0	0	0	3	0	0	0	0	3.6	0
	25	Walkthroughs	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4.0	0
	26	System-improvement	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0	0

The results in the third dimension (Policy Insights and Spillovers) had a growing number of "do not apply", which is representative of the initial maturity level of the process. The negative evaluations are still negligible and the results from the previous section on "follow-up and integration session" showed that two DAs were able to produce the "**initial concrete measures"**, revealing that some insights and recommendations were made.

The last dimension (Administrative, Review, and Overview) had the largest number of "do not apply", These results were expected since some of the later criteria represent challenges to the modeling process and rarely would have a positive evaluation if the process is at its early stages of development. In addition, the positive results were still much higher than the negative.

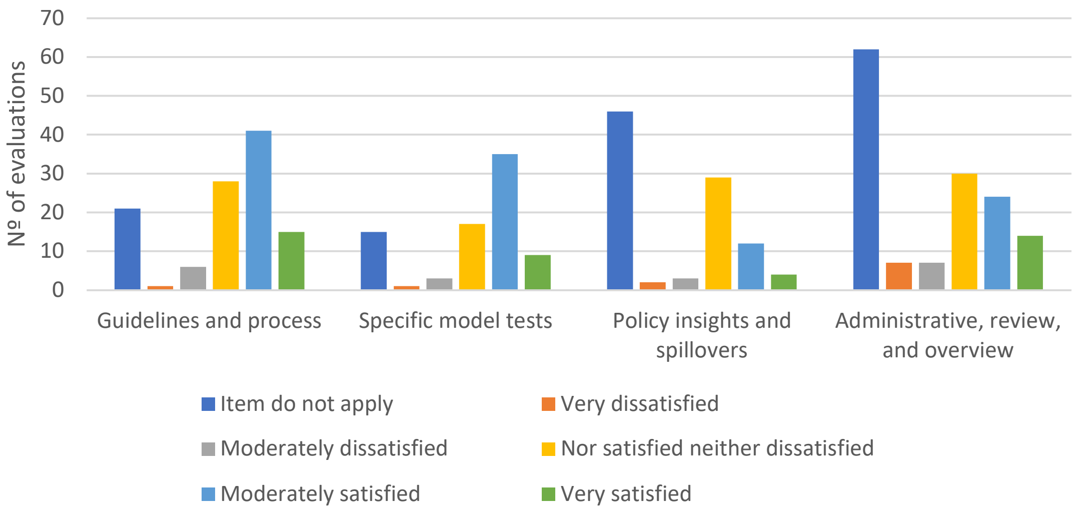

Figure 49: Aggregate distribution of evaluations by dimension of the workshops organized on the Causal Loop Diagrams process.

Dimension 1: Guidelines and process

This dimension embraces indicators 1 to 7 (Table 34) and has 85 comments (Appendix 3). The average results varied from 3.3 (boundary adequacy) to 4.1 (Meaningfulness of the process). This dimension represents the initial steps of the modeling process and it had many more positive evaluations (56) when compared to negative ones (7). This is a very positive result because the time dedicated to model preparation and group model construction was short (as commented by several participants of the exercise).

Dimension 2: Specific model tests

This dimension embraces indicators 8 to 12 (Table 34) and has 60 comments. The second dimension (Specific Model Tests) had more positive results (44) than the sum of all other categories of answers (36). The average results varied from 3 (extreme conditions) to 4.1 (boundary adequacy). The first was known to be controversial due to its relations to a numerical model and it was expected to be problematic. The second is very important as it discusses the limitations of the model, and how the DAs evaluate it.

Dimension 3: Policy insights and spillovers

This dimension embraces indicators 13 to 18 (Table 34) and has 62 comments. The average result varied from 2.9 (insight generation capacity) to 3.8 in learning. Despite the DAs low evaluation of the insight generation capacity (which was whatsoever expected in this stage of the modeling process), they did produce insights, described in the section "follow-up and integration section". Most of the respondents about learning related to a positive experience but are still confused with the stakeholders' participation.

Dimension 4: Administrative, review, and overview

This dimension embraces indicators 19 to 26 (Table 34) and has 87 comments. The average values vary from 2.5 (Replicability) to 3.9 (time and cost of intervention). Replicability of a group model building is virtually impossible because the model is a product of a group discussion. The methods to conduct the sessions are replicable, are described in the cited literature, and are briefly described in the methods section of this chapter. Regarding the time and cost of the exercise, most people were satisfied with the process and the time involved.

In conclusion, the four dimensions of the evaluation process (Guidelines and Processes, Specific Model Tests, Policy Insights and Spillovers, and Administrative, Review, and Overview) captured the variations in the DA view regarding the criteria used for the validation, with a great prevalence of positive aspects regarding the process, what reveals the CLDs are a robust tool to proceed in Marine SABREs analysis.

10.5 Conclusions of the CLD Building Process

This section presents and briefly discusses the methods and processes conducted with the DAs to produce the new set of causal loop diagrams to be used in Marine SABRES analysis from now on.

By the results obtained, we claim the modeling process was a productive exercise, and despite the limitations of the method, human resources, time, etc., the product represents the **major components of the system**, as viewed by the DAs scientists.

Four categories of results were presented, namely: a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol. **These four types of results complement and integrate each other and are integral parts of the baseline assessment of the Social-Ecological models as required by the DoA of Marine SABRES to be present in D4.1.**

In addition, by conducting the process, we conclude the present **baseline assessment of the Social-Ecological models** is**:**

- **Meaningful**, as stated prolifically by the DAs (Appendix 2 and 3).

- **Robust**, as revealed by the great prevalence of positive aspects and very low rejection of the present model, captured by the validation process (Figure 48-49, Table 34 and Appendix 3).

- **Integral**, the models of all DAs (especially the Artic DA and Macaronesia which is a novelty) are integral representations of the issues and regions/countries of this DA, as requested by the DoA.

As such the results from the present section are considered adequate to be used for the upcoming tasks of the project as demonstrated below.

10.6 Recommendations on the way forward

WP4 recommends all interested parts of Marine SABRES to use the results presented in this section of the **baseline assessment of the Social-Ecological models,** especially to develop the next steps from WP5 (Table 35).

WP4 is currently using the results of Section 10 to develop the storylines and pathways (in short, the "storylines") that will lead to the desired scenarios mapped in D5.1. This is currently being done using a series of workshops and will be described in D4.2. These storylines will complement the baseline assessment of the Social-Ecological models presented here and will be available for the other WPs probably in late July 2024. These storylines represent the first stage of the "options and pathways report" (D4.2).

Table 35:Recommendations for users of the results here baseline assessment of the Social-Ecological models

Item	Description	User	Task
Baseline assessment of the Social-Ecological models plus the storylines and pathways described in D4.2	The collection of a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol	WP4 CEFAS	D4.2 storylines from each DA showing the pathways to transformations
Baseline assessment of the Social-Ecological models plus the storylines and pathways described in D4.2	The collection of a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol, plus the storylines and pathways described in D4.2	WP5.2 WP4	An ex-ante impact evaluation of options for development will be performed, including an economic assessment of the proposed actions (D5.2)

Parts b) the CLD model; c) the follow-up and integration questionnaire added by the description of pathways and storylines.	Probably these two parts are more relevant for the graph exercise of modeling and testing the trade-offs. As the possible elements to change in the system need to be mapped previously, the contributions from enablers and barriers seen by the DAs (described in the follow-up and integration questionnaire, plus storylines) will be useful.	WP5.3, WP5.2 WP4	Conflicts and trade-offs emerging from future scenarios (based on Task 5.1 and Task 5.2) will be identified. The behavioral subsystems of the SES will be used to identify and quantify in euro, the incentives necessary to promote individual and collective action to resolve these conflicts and settle these trade-offs toward sustainability and resilience goals (D5.3).
Baseline assessment of the Social-Ecological models plus the storylines and pathways described in D4.2	Some governance arrangements for ABNJ were already pointed out and commented on by two DAs (Arctic and Macaronesia) with the possibility of starting a “regional dialogue for the corridor” being considered a great positive result by the Macaronesia DA. As this and other relevant topics are described and commented on in a) the problem articulation session description, b) the CLD model; c) the follow-up and integration questionnaire; and d) the answers to the validation protocol, they are considered relevant as the complementary information from the storylines in D4.2.	WP5.4 WP5.3, WP5.2 WP4	Recommendations for governance approaches required to meet the sustainability and resilience objective of the DA scenarios will be developed. This will include an assessment of innovative governance arrangements and capacity needed for Area Based Management Tools including MPAs in ABNJ. The recommendations will promote effective and legitimate governance arrangements for EBM, including challenges and opportunities for cross-sectoral and transboundary cooperation and overcoming implementation drift. (D5.4)
Baseline assessment of the Social-Ecological models	As the report brings invaluable knowledge about the tasks in the testing of the simple SES and building the new CLDs, it can be useful for the refinement phase of the simple SES approach.	WP3	Once the proposed Simple SES has been tested in WP4 by the DA and their findings collated in a second iteration, WP3 will amend the Simple SES for delivery into WP6 (D3.3).
Baseline assessment of the Social-Ecological models plus the storylines and pathways described in D4.2	The results in the present baseline assessment of the Social-Ecological models can be used to inform the DA-SG, and specifically test their acceptance (cultural, behavioral, economic) of the loops in their systems, and of the storylines and scenarios currently being developed.	WP2	Stakeholders from the DA SGs will test the initial, as well as the semi-final, outcomes of the other WPs, and will discuss the test results in a series of workshops. The appraisal of the scenarios and costed options by each DA-SG using a common format.

Task 5.2 can build the ex-ante economic evaluation, using the results in the present baseline assessment of the Social-Ecological models and those that will be delivered in the storylines (representing the "current and the alternative conditions" (optional pathways) in the system), that each DA could take to deal with their challenges and lead to their desired futures. Furthermore, **the specific knowledge described in D4.1 and D4.2 will show trade-offs, and synergies of conservation to biodiversity and other developments, including some barriers and enablers as seen by the DAs.** Lastly, the CLD produced in this section show the most relevant economic sectors in each DA. This will specifically feed into the requirement of an "understanding the impact of responses on sectors and specific business types (tourism operators, commercial fishers, etc.).", as this is the basis for the economic evaluation in T5.2 of options for development.

Taks 5.3 will investigate the Societal behavioral influence on sustainability issues. **The models present in this section**, accompanied by their descriptions and the storylines that lead to relevant and desired futures, **are the basis where task 5.3 can feed** to draw the understanding of the system and where behavioral change is needed. Both the CLDs and the storylines (to be developed in D4.2) derived from them will show the "conflicts and trade-offs emerging from scenarios" and they can show the relevant points in the system necessary to promote collective actions towards desired scenarios, as well as reveal where the limitations of these collective actions are. The present CLD is the basis where the "modeling changes in the value of links between subsystems" must be done to understand the issues around sustainability and resilience goals (D5.3).

Task 5.4 will need to build "recommendations for governance approaches required to meet the sustainability and resilience objective of the DA scenarios, and an assessment of innovative governance arrangements and capacity needed for Area Based Management Tools including MPAs in ABNJ". **The results in the present baseline assessment** of the Social-Ecological models **are crucial for T5.4 as they show the validated by DAs understanding of the system** and they fulfill the legitimacy requirement for the governance arrangements by the DoA. In addition, they show ideas of governance measures, internally to the DAs but also in ABNJ (clearly in Tuscany and Macaronesia) that should be explored in terms of possibilities of new arrangements and its limits to the current conditions

Apart from tasks in WP5, **the results in the present baseline assessment of the Social-Ecological models can contribute to tasks in WP2 and WP3:** Task 3.3 embraces the refinement of the simple SES tool, and therefore the content in the present deliverable can provide an invaluable contribution to this task, as it brings many comments from the DAs regarding the whole experience of testing both the sSES approach and the new CLD tool.

As **Task 2.3 is conducive to the appraisal and feedback of the WP outcomes, the workshops in this task can use the loop leading the variables of interest (PESTLE**) in each DA to understand their acceptance in cultural, economic, behavioral, and other terms. This approach and results will directly contribute to the requirements of "Stakeholders from the DA SGs will test the initial, as well as the semi-final, outcomes of the other WPs, and will discuss the test results in a series of workshops (M2.10-M2.17). Appraisal of the scenarios and costed options (see T5.1, T5.2) by each DA-SG using a common format." As the present baseline assessment of the Social-Ecological models and the D4.2 show the trade-offs, enablers and barriers, and major options towards sustainability, they can be directly involved with the "A common survey methodology will be applied in each DA including scoring criteria on efficiency effectiveness and equity and the major trade-offs involved. Barriers and enablers towards the chosen pathways will be identified by each DA-SG and recommendations for implementation will be developed (D2.3)" requirements.

With these exercises, we have completed the baseline system function and the qualitative assessment for the SES models from each DA. This should be validated by stakeholders in WP2 and WP6 (Task 2.3 and T6.2), to ensure that they captured the major components of system function and are sufficiently robust to support decisions. **This Deliverable 4.1 constitutes a starting point for further development and trialling in each DA. The conclusions obtained in this study are to be used by WP3 in a second iteration of the sSES being carried out by the DAs, in order to refine and improve the sSES approach and its paper (or electronic) guidance, as necessary, to be used in other locations as a stand-alone method. Then results in Section 10 should be used by WP5 to proceed with their economic, behavioral and governance analysis of the DAs.** The results of these applications with the three DAs will be used when creating targeted solutions within WP5 and eventually be upscaled in WP6.

We suggest that results in this Deliverable can be already used for an initial appraisal of the situation in the DAs as some exploratory navigation of the systemic properties of each variable might be profitable for management, taking into account the initial maturity stage of the model in Section 10. In addition, **we recommend using the results of this baseline assessment, especially the loops and the variables of interest** (described in Section 10) **to deepen the systems analysis using stakeholder inputs as complementary knowledge.** The fact that distinct groups often see the systems in different ways and frame them in different model structures is trivial in systems analysis. Therefore, it is emphasized that there is little sense in comparing the structure of the CLDs presented in this deliverable with a possible structure made by stakeholders. Therefore, **it is suggested that the CLDs be used to communicate the expert views of the system structure and then complemented punctually inside a small set of loop analyses, by specific knowledge from the stakeholders' groups.** This approach can complement the results presented here, ensuring they capture the major components of system functioning and are sufficiently robust to support decisions (Task 2.3). Initial concrete management measures are identified in the DA descriptions, these are the starting point for further concrete measures development and trialling in each DA.

REFERENCES

Ackoff, R. L. (1956). The Development of Operations Research as a Science. Operations Research, 4(3), 265-295.

Andersen, D. L., Luna-Reyes, L. F., Diker, V. G., Black, L., Rich, E., & Andersen, D. F. (2012). The disconfirmatory interview as a strategy for the assessment of system dynamics models. System Dynamics Review, 28(3), 255-275. https://doi.org/10.1002/sdr.1479

Anon 2023a. Report of the Coastal States Working Group on the Distribution of Northeast Atlantic Mackerel. October 2023. 115pp.

Anon., 2023b. Report of the 2023 Coastal States Working Group on the distribution of blue whiting (Micromesistius poutassou) in the Northeast Atlantic. October 2023. 82pp.

Anon., 2023c. Report of the Coastal States Working Group on the distribution of Norwegian spring spawning herring in the North-East Atlantic and the Barents Sea. October 2023. 81pp.

Atkins, J.P., Smith, G., 2023. Indicators for the Simple SES approach Briefing Paper, Deliverable 3.2. Marine SABRES, European Union's Horizon Europe research and innovation programme under grant agreement no. 101058956. and the UKRI Project Numbers 10050525 and 10040244

Atkins, J.P., Burdon, D., Elliott, M., Gregory, A.J., 2011. Management of the Marine Environment: Integrating Ecosystem Services and Societal Benefits with the DPSIR Framework in a Systems Approach. Marine Pollution Bulletin, 62(2): 215–226.

Balci, O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. Annals of Operations Research, 53(1), 121-173. https://doi.org/10.1007/BF02136828

Barcelona, A., C. Oldham, J. Colomer, J. Garcia-Orellana, T. Serra, 2021. Particle Capture by Seagrass Canopies under an Oscillatory Flow. Coastal Engineering 169:103972. doi: 10.1016/j.coastaleng.2021.103972.

Barlas, Y. (1989). Multiple tests for validation of system dynamics type of simulation models. European Journal of Operational Research, 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. System Dynamics Review, 12(3), 183-210. https://doi.org/10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Barlas, Y., & Carpenter, S. (1990). Philosophical roots of model validation: Two paradigms. System Dynamics Review, 6(2), 148-166.

Bell, S.; S. Morse, 2013. Rich Pictures: a means to explore the 'Sustainable Mind'?. Sustainable Development. 21. 10.1002/sd.497.

Bremner, J., & Oliveira, B. (2024). Scenarios Briefing Paper. Deliverable 3.2 Marine SABRES, European Union's Horizon European research and innovation programme under grant agreement no. 101058956. and the UKRI Project Number 10050525.

Bremner, J., Oliveira, B., Mynott, F., Pinnegar, J. K., & Huertas, C. (2024). Scenarios to reach biodiversity, sustainability and climate targets: Developing pathways to transformation. Deliverable 5.1. Marine SABRES, European Union's Horizon European research and innovation programme under grant agreement no. 101058956. and the UKRI Project Number 10050525.

CAFF, 2012. The Arctic Species Trend Index 2011. Available from: https://oaarchive.arctic-council.org/server/api/core/bitstreams/49c44140-4a7e-44b8-ace1-b317ff625b6b/content (last accessed 15 Jan 2024).

Checkland, P. B. (1989). Soft Systems Methodology*. Human Systems Management, 8(4), 273-289. https://doi.org/10.3233/HSM-1989-8405

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

Chowdhury, R., Gregory, A., & Queah, M. (2023). Creative and flexible deployment of systems methodologies for child rights and child protection through Holistic Flexibility. Systems Research and Behavioral Science, 40(4), 654-670. https://doi.org/10.1002/sres.2955

Churchman, C. W., Ackoff, R. L., & Arnoff, E. L. (1957). Introduction to operations research.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. Psychological Assessment, 31(12), 1412-1427. https://doi.org/10.1037/pas0000626

Crabolu, G., Font, X., & Eker, S. (2023). Evaluating policy complexity with causal loop diagrams. Annals of tourism research, 100, 103572.

DeFries, R., & Nagendra, H. (2017). Ecosystem management as a wicked problem. Science, 356(6335), 265-270.

Douglas, M., & Wildavsky, A. (1983). Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers.

Duarte, C. M., J. J. Middelburg, N. Caraco, 2005. Major Role of Marine Vegetation on the Oceanic Carbon Cycle. Biogeosciences 2(1):1–8. doi: 10.5194/bg-2-1-2005.

Duffy, J. E., L. Benedetti-Cecchi, J. Trinanes, F.E. Muller-Karger, R. Ambo-Rappe, C. Boström, A.H. Buschmann, J. Byrnes, R.G. Coles, J. Creed, L.C. Cullen-Unsworth, G. Diaz-Pulido, C. M. Duarte, G.J. Edgar, M. Fortes, G. Goni, C. Hu, X. Huang, C.L. Hurd, C. Johnson, B. Konar, D. Krause-Jensen, K. Krumhansl, P. Macreadie, H. Marsh, L.J. McKenzie, N. Mieszkowska, P. Miloslavich, E. Montes, M. Nakaoka, K. Magnus Norderhaug, L.M. Norlund, R. J. Orth, A. Prathep, N.F. Putman, J. Samper-Villarreal, E.A. Serrao, F. Short, I. Sousa Pinto, P. Steinberg, R. Stuart-Smith, R.K.F.F. Unsworth, M. van Keulen, B.I. van Tussenbroek, M. Wang, M. Waycott, L.V. Weatherdon, T. Wernberg, S.M. Yaakub. 2019. Toward a Coordinated Global Observing System for Seagrasses and Marine Macroalgae. Frontiers in Marine Science 6(JUL). doi: 10.3389/fmars.2019.00317.

Elliott, M., O'Higgins, T.G., 2020. From the DPSIR, the D(A)PSI(W)R(M) emerges… a butterfly. In: Ecosystem-Based Management and Ecosystem Services: Theory, tools and practice. T.G., O'Higgins, M Lago & T.H. DeWitt (Eds.). Springer, Amsterdam. ISBN 978-3-030-45842-3, ISBN 978-3-030-45843-0 (eBook); https://doi.org/10.1007/978-3-030-45843-0; p61-86

Elliott, M., Borja, A., Cormier, R., 2020. Managing marine resources sustainably: a proposed integrated systems analysis approach. Ocean & Coastal Management, 197, 105315, https://doi.org/10.1016/j.ocecoaman.2020.105315

EU Commission, 2021. EU and Greenland reach agreement on new fisheries partnership—European Commission. https://oceans-and-fisheries.ec.europa.eu/news/eu-and-greenland-reach-agreement-new-fisheries-partnership-2021-01-08_en

Florencio, M., Patiño, J., Nogué, S., Traveset, A., Borges, P. A., Schaefer, H., ... & Santos, A. (2021). Macaronesia as a fruitful arena for ecology, evolution, and conservation biology. Frontiers in Ecology and Evolution, 9, 718169).

Forrester, J. W. (1961). Industrial dynamics. MIT Press: Cambridge, MA.

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

Forrester, J. W., Low, G. W., & Mass, N. J. (1974). The debate on World Dynamics: A response to Nordhaus. Policy Sciences, 5(2), 169-190. https://doi.org/10.1007/BF00148039

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. Études et Recherches Sur Les Systèmes Agraires et Le Développement, 169.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. Futures, 25(7), 739-755. https://doi.org/10.1016/0016-3287(93)90022-L

Gercek, E., Gambino, M., & Malvarosa, L. (2022). Understanding EU Fisheries Management Dynamics by Engaging Stakeholders through Online Group Model-Building. Sustainability, 14(23), 15862.

Government of the Faroe Islands, 2024. Available from: https://www.government.fo [last accessed: 24.01.2024]

Granata, TC, T. Serra, J. Colomer, X. Casamitjana, CM Duarte, E. Garcia, 2001. Flow and Particle Distributions in a Nearshore Seagrass Meadow before and after a Storm. Marine Ecology Progress Series 218:95–106. doi: 10.3354/meps218095.

Gregory, A.J., Atkins, J.P., Smith, G., Elliott, M., 2023. Simple Social-Ecological Systems Guidance, Deliverable 3.1. Marine SABRES, European Union's Horizon Europe research and innovation programme under grant agreement no. 101058956. and the UKRI Project Number 10050525

Grønlands Statistik, 2023. Greenland in Figures—2023. https://stat.gl/dialog/topmain.asp?lang=en&subject=&sc=GF

Haynes, M. G. (1995). Soft Systems Methodology. En K. Ellis, A. Gregory, B. R. Mears-Young, & G. Ragsdell (Eds.), Critical Issues in Systems Theory and Practice (pp. 251-257). Springer US. https://doi.org/10.1007/978-1-4757-9883-8_35

Hummel, H. Segeren, R. Sturm, B. Thijssen, B., 2023. Simple SES design brief, Deliverable 2.1. European Union's Horizon Europe research and innovation programme grant agreement no. 101058956.

Iceland Responsible Fisheries, 2021. Icelandic Fisheries. Responsible Fisheries.is. icelandic-fisheries-press-kit-mai-2021-enska.pdf (responsiblefisheries.is)

ICES, 2022a. Icelandic Waters ecoregion – Ecosystem overview (Report of the ICES Advisory Committee) [Report]. ICES Advice: Ecosystem Overviews. https://doi.org/10.17895/ices.advice.21731663.v1

ICES, 2022b. Norwegian Sea Ecoregion—Ecosystem Overview [Report]. ICES Advice: Ecosystem Overviews. https://doi.org/10.17895/ices.advice.21731726

ICES, 2023a. Working Group on Widely Distributed Stocks (WGWIDE). ICES Scientific Reports. 5:82. 980 pp. https://doi.org/10.17895/ices.pub.24025482

ICES, 2023b. Working Group on the Integrated Assessments of the Norwegian Sea (WGINOR, outputs from 2022 meeting). ICES Scientific Reports. 5:15. 57 pp. https://doi.org/10.17895/ices.pub.22110260

ICES, 2023c. Faroes Ecoregion—Ecosystem Overview [Report]. ICES Advice: Ecosystem Overviews. https://doi.org/10.17895/ices.advice.24711000.v1

ICES, 2023d. Greenland Sea ecoregion – Ecosystem overview [Report]. ICES Advice: Ecosystem Overviews. https://doi.org/10.17895/ices.advice.22664881.v1

Jebb, A. T., Ng, V., & Tay, L. (2021). A Review of Key Likert Scale Development Advances: 1995–2019. Frontiers in Psychology, 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.637547

Kokorsch, M., Benediktsson, K., 2018. Prosper or perish? The development of Icelandic fishing villages after the privatisation of fishing rights. Maritime Studies, 17: 69-83. Doi: 10.1186/s40152-018-0089-5.

Käll, S., Crona, B., Van Holt, T., & Daw, T. M. (2022). From good intentions to unexpected results—A cross-scale analysis of a fishery improvement project within the Indonesian blue swimming crab. Maritime Studies, 21(4), 587-607. https://doi.org/10.1007/s40152-022-00285-y

Kuhn, T. S. (1962). The structure of scientific revolutions (Vol. 962). University of Chicago press Chicago.

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. Proceedings of the System Dynamics Society International Conference, 67-77.

Latour, B. (2013). Laboratory life: The construction of scientific facts. Princeton university press.

Lemke, J., & Łatuszyńska, M. (2013). Validation of System Dynamics Models – a Case Study. Journal of Entrepreneurship, Management and Innovation, 9(2), Article 2. https://doi.org/10.7341/2013923

Marine Conservation Institute, 2024. https://mpatlas.org/countries/GRL/ [last accessed 9. February 2024]

Marine Research Faroes, 2023. https://www.hav.fo/ [last accessed: 15. February 2024]

McGlathery, KJ, K. Sundbäck, IC Anderson, 2007. Eutrophication in Shallow Coastal Bays and Lagoons: The Role of Plants in the Coastal Filter. Marine Ecology Progress Series 348:1–18. doi: 10.3354/meps07132.

Meadows, D. (1980). The Unavoidable a Priori. En J. Randers (Ed.), Elements of the system dynamics method. Pegasus Communication.

Meadows, D. H. (2008). Thinking in systems: A primer. chelsea green publishing.

MFRI, 2023. Capelin advice. Available from: https://www.hafogvatn.is/static/extras/images/31-capelin-autumn1408221.pdf (last accesses 15 Jan 2024).

Montefalcone, M., P. Vassallo, G. Gatti, V. Parravicini, C. Paoli, C. Morri, C. Nike Bianchi, 2015. The Exergy of a Phase Shift: Ecosystem Functioning Loss in Seagrass Meadows of the Mediterranean Sea. Estuarine, Coastal and Shelf Science 156:186–94. doi: 10.1016/j.ecss.2014.12.001.

Morecroft, J. D. W., & Sterman, J. D. (1994). Modelling for Learning Organizations.

Mukherjee, N., Hugé, J., Sutherland, W.J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., Koedam, N., 2015. The Delphi technique in ecology and biological conservation: applications and guidelines. Methods Ecol Evol, 6: p. 1097–1109. https://doi.org/10.1111/2041-210X.12387

NAMMCO, 2019. Estimates of the Abundance of Cetaceans in the Central North Atlantic based on NASS Icelandic and Faroese Shipboard Surveys Conducted in 2015. https://doi.org/10.7557/3.4941

Ney, S. (2012). Resolving messy policy problems: Handling conflict in environmental, transport, health and ageing policy. Routledge.

Nordhaus, W. D. (1973). World Dynamics: Measurement Without Data. The Economic Journal, 83(332), 1156-1183. https://doi.org/10.2307/2230846

Nordisk Ministerråd, Høst, J., Christiansen, J., 2018. Nordic fisheries in transition: – Future challenges to management and recruitment (2018:545; p. 126). Nordisk Ministerråd. https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-5355

Oliveira, B. (2022). Ecosystem-based management of social-ecological systems: A required perspective and agenda. En Oliveira, B; Pardo, J. & Turra, A. (Orgs) Challenges in ocean governance in the views of early career scientists: Achievements of the São Paulo School of Advanced Science on Ocean. S (p. 235). Instituto de Estudos Avançados da Universidade de São Paulo.

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

Overmars, K. P., de Groot, W. T., & Huigen, M. G. (2007). Comparing inductive and deductive modelling of land use decisions: Principles, a model and an illustration from the Philippines. Human Ecology, 35, 439-452.

Perry, A. H., 2000. Impacts of Climate Change on Tourism in the Mediterranean: Adaptive Responses. SSRN Electronic Journal.

Qudrat-Ullah, H. (2012). On the validation of system dynamics type simulation models. Telecommunication Systems, 51(2), 159-166. https://doi.org/10.1007/s11235-011-9425-4

Qudrat-Ullah, H., & Seong, B. S. (2010). How to do structural validity of a system dynamics type simulation model: The case of an energy policy model. Energy Policy, 38(5), 2216-2224. https://doi.org/10.1016/j.enpol.2009.12.009

Radzicki, M. J. (1990). Methodologia oeconomiae et systematis dynamis. System Dynamics Review, 6(2), 123-147.

Reisman, A., & Oral, M. (2005). Soft Systems Methodology: A Context within a 50-Year Retrospective of OR/MS. Interfaces, 35(2), 164-178.

Richardson, G. P. (1986). Problems with causal-loop diagrams. System Dynamics Review, 2(2), 158-170. https://doi.org/10.1002/sdr.4260020207

Richardson, G. P. (1997). Problems in causal loop diagrams revisited. System Dynamics Review, 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. System Dynamics Review. https://onlinelibrary.wiley.com/doi/abs/10.1002/sdr.4260100204

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), 155-169. https://doi.org/10.1007/BF01405730

Rutty, M., D. Scott, 2010. Will the Mediterranean Become "Too Hot" for Tourism? A Reassessment. Tourism and Hospitality Planning & Development 7:267–281.

Schwaninger, M., & Groesser, S. (2020). System Dynamics Modelling: Validation for Quality Assurance. En B. Dangerfield (Ed.), System Dynamics: Theory and Applications (pp. 119-138). Springer US. https://doi.org/10.1007/978-1-4939-8790-0_540

Schwarz, M., & Thompson, M. (1990). Divided We Stand: Redefining Politics, Technology, and Social Choice. University of Pennsylvania Press.

Scolobig, A., Thompson, M., & Linnerooth-Bayer, J. (2016). Compromise not consensus: Designing a participatory process for landslide risk mitigation. Natural Hazards, 81(1), 45-68. https://doi.org/10.1007/s11069-015-2078-y

SFS, 2023. Losun koldvísýrings. Available from: https://radarinn.is/Umhverfismal/LosunCO2 (last accessed 15 Jan 2024).

Smith, G. Gregory, A. Atkins, J. Elliott, M. (2023) Review of the Literature on Social-Ecological Systems. Deliverable 3.1. Marine SABRES, European Union's Horizon Europe research and innovation programme under grant agreement no. 101058956. and the UKRI Project Number 10050525

Statistics Faroe Islands, 2024. Available from: https://hagstova.fo/fo [last accessed: 24.01.2024]

Statistics Greenland, 2024. Available from: https://bank.stat.gl/pxweb/en/ [last accessed: 15.02.2024]

Statistics Iceland, 2023. Available from: https://www.statice.is/ (last accessed 15 Jan 2024).

Sterman. (2000). Business Dynamics: Systems Thinking and Modelling for a Complex World with CD-ROM.

Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. System Dynamics Review, 18(4), 501-531. https://doi.org/10.1002/sdr.261

Telesca, L., A. Belluscio, A. Criscoli, G. Ardizzone, E.T. Apostolaki, S. Fraschetti, M. Gristina, L. Knittweis, C.S. Martin, G. Pergent, A. Alagna, F. Badalamenti, G. Garofalo, V.

Gerakaris, M.L. Pace, C. Pergent-Martini, M. Salomidi, 2015. Seagrass Meadows (Posidonia Oceanica) Distribution and Trajectories of Change. Scientific Reports 5(1):12505. doi: 10.1038/srep12505.

The Danish Parliament, 2021, October 27. Greenland and the Faroe Islands / The Danish Parliament. https://www.thedanishparliament.dk/en/eu-information-centre/greenland-and-the-faroe-islands

The Ministry of Fishing and Hunting, 2021. Sectoral Policy Programme Document 2021 – 20: https://naalakkersuisut.gl/-/media/departementer/fiskeri_fangst/publikationer/sectoral-policy-programme-document-2021-2024---eng.pdf . [last

The Faroe Islands Fisheries Inspection, 2014. https://www.vorn.fo/ [last accessed: 15. February 2024]

The Nordic Council, 2024. https://www.norden.org/en/nordic-council [last accessed 9. February 2024]

Thompson, M. (1997). Cultural theory and integrated assessment. Environmental Modelling & Assessment, 2(3), Article 3. https://doi.org/10.1023/A:1019065412191

Thompson, M., Ellis, R. J., Wildavsky, A., & Wildavsky, M. (1990). Cultural Theory. Avalon Publishing.

Toulmin, S. (1977). From Form to Function: Philosophy and History of Science in the 1950s and Now. Daedalus, 106(3), 143-162.

Unsworth, R.K.F., L. Mtwana Nordlund, L.C. Cullen-Unsworth, 2019. Seagrass Meadows Support Global Fisheries Production. Conservation Letters 12(1):e12566. doi: 10.1111/conl.12566.

Van den Belt, M. (2004). Mediated modelling: A system dynamics approach to environmental consensus building. Island press.

Verweij, M., Douglas, M., Ellis, R., Engel, C., Hendriks, F., Lohmann, S., Ney, S., Rayner, S., & Thompson, M. (2006). Clumsy Solutions for a Complex World: The Case of Climate Change. Public Administration, 84(4), 817-843. https://doi.org/10.1111/j.1540-8159.2005.09566.x-i1

Vigliano Relva J, Jung J, 2021. Through the Eyes of Another: Using a Narrative Lens to Navigate Complex Social-Ecological Systems and to Embrace Multiple Ways of Knowing. Front. Mar. Sci. 8:678796; https://doi.org/10.3389/fmars.2021.678796

Waycott, M., C. M. Duarte, T. J. B. Carruthers, R. J. Orth, W. C. Dennison, S. Olyarnik, A. Calladine, J. W. Fourqurean, K. L. Heck, A. R. Hughes, G. A. Kendrick, W. J. Kenworthy, F. T. Short, S. L. Williams, 2009. Accelerating Loss of Seagrasses across the Globe Threatens Coastal Ecosystems. Proceedings of the National Academy of Sciences 106(30):12377–81. doi: 10.1073/pnas.0905620106.

Appendix 1: Validation protocol detailed description

A1 Introduction

Two types of models form the main tools in system dynamics: quantitative stock-and-flow models and qualitative Causal Loop Diagrams (CLDs) (Sterman, 2000). The present piece focuses on the CLDs due to their relatively reduced level of complexity and coherence with the model we focus on to work, the product of the sSES approach. In addition, these system dynamics models represent a set of causality models most people are not very familiar with.

Causality models represent a form of a theory of how a system works. It describes the connections of the elements of the system in such a way that one can create an understanding of the system by understanding the causalities described in the model. On the other hand, statistical models are based on ideas of correlation between variables in the system that can be used to forecast or predict the behaviour of a system, preferably inside the same parameter range to which the correlation was observed. However, they cannot produce the same explication feature offered by a causal model (Barlas and Carpenter, 1990). This movement can be understood as an analogy of the evolutionary perspective of the 1950-1970s scientific transition in which “the positivists' earlier preoccupation with "prediction," which they had regarded as the key evidence of scientific \`knowledge,´ was being supplemented by a concern with \`explanation,´ regarded as the core of scientific \`understanding´ (Radzicki, 1990; Toulmin, 1977). This methodological/philosophical dichotomy has been tested recently and showed that the predictive capacity of both causal (deductive) and statistical (inductive) models are equivalent with the advantage of causal models providing additional explicative power to the results (Overmars et al., 2007).

A brief review of the underlying philosophies concerning scientific theory development will point out that the dichotomy in these branches is deeper than purely methodological. Authors claim there are philosophical schools that might justify the division shown above and additionally underpin the assumptions of validation of theory/models, namely: the empiricist/reductionist school and the relativist/holistic school (Barlas and Carpenter, 1990). This division into two main branches in the philosophy of science was corroborated in Economics studies, although with distinct names, where the empiricist/reductionist was called neoclassical (logical empiricism) and relativist/holistic named institutionalists (pragmatic instrumentalism) (Radzicki, 1990).

The empiricist/reductionist school is a perspective rooted in Kant's epistemology and based on ideas of knowledge being entirely objective, ahistorical, asocial, and acultural, to which an absolute truth can be reached independently of human values and belief (Barlas and Carpenter, 1990). This school, posteriorly discussed by Russel, the early Wittgenstein, and the contributions of the Vienna Circle, evolved to the logical empiricism of the early 20th century, which focused on the reduction of scientific statements to the criteria of being validated by the direct observational statements, with great rigor in the meaning of each statement, and avoiding ambiguities, vagueness, and inconsistencies. Popper collaborated with this school in his early days, by advocating the criteria of falsifiability (instead of verifiability) to which a theory gains trust as much as failed attempts to prove it wrong accumulate. From this falsifiability idea comes the statement that every piece of knowledge is fallible because its status of valid is always provisional due to the possibility of being falsified by new evidence. One of the main critiques here comes from the Kuhnian perspective of scientific knowledge being biased by the Normal ruling paradigm, and consequently heavily historically and socially influenced (Kuhn, 1962).

The relativist/holistic school is a perspective rooted in Hegel´s epistemology, where scientific ideas are byproducts of an Age, influenced directly or indirectly by the social foundations in which it was created. Knowledge is, thus, socially, historically, and culturally dependent (Latour, 2013), therefore there cannot be a neutral foundation, and a pure objective verification is not possible (Barlas and Carpenter, 1990). From this school, the idea of pure knowledge, independent of social and historical processes was abandoned towards interdisciplinary more flexible ideas, where the absolute truth out of formal rigor, opens space for more functional perspectives: "The academic `soundness´ required rigor of a kind that, in these grey interdisciplinary areas was simply not there to be had" (Toulmin, 1977).

This relativist/holist perspective is convergent with other theories in complexity science that represent the vanguard of interdisciplinary knowledge when tackling problems of the 21st century. These recent approaches understand the part of complex problems regarding society as wicked or messy (instead of tame) (DeFries and Nagendra, 2017; Ney, 2012; Rittel and Webber, 1973; Verweij et al., 2006), to which policy problems are socially defined, and therefore dependent on the plurality of views society produces (see plural views below). Moreover, when considering a plural society, public goods are in dispute, meaning public policies cannot be correct or false, they are always dependent on each social group these policies are representing. Public policies cannot propose an "optimal solution" since what is optimal for one group, might be the obliteration of others. This perspective sees the boundaries of the pressing problem as becoming less clear-cut as the connectivity of global society increases, and thus, far more dependent on framing, debate, and controversies, byproducts of a plural society. In short:

"…in a pluralistic society, there is nothing like the undisputable public good; there is no objective definition of equity; policies that respond to social problems cannot be meaningfully correct or false; and it makes no sense to talk about "optimal solutions" to social problems unless severe qualifications are imposed first." (Rittel and Webber, 1973)

A plural view of society, as described by the theory of plural rationalities (Douglas and Wildavsky, 1983), understands the scientific work as biased by worldviews, despite the claims of neutrality of the scientific community. The basic assumption is that social relations provide the individual with normative and cognitive tools to understand the world (shared values and beliefs). Here, politics, decision-making, technology, and social choice are understood as dependent on cultural backgrounds, described by shared values and beliefs (or worldviews), to which a typology framework (Schwarz and Thompson, 1990; Thompson, 1997) is proposed. With this framework, conflicting perspectives about the pressing problems can be understood and managed by a conflict-reducing heuristic (Ney, 2012; Oliveira, 2022; Scolobig et al., 2016).

These interdisciplinary approaches are complemented by the perspective of Post-Normal science (Funtowicz and Ravetz, 1997) since it provides an understanding of science with greater openness to democratic participation. In post-normal science, “facts are uncertain, values in dispute, stakes high and decisions urgent” (Funtowicz and Ravetz, 1993). This perspective enhances the idea of stakeholder consultation to broader and deeper participation, expanding the conception of stakeholders, usually restricted to the scientific community plus relevant decision-makers, to the broad community (expanded peer community) based on the justification of shared risks of the globalized civilization.

The relevance of this discussion to the validation of system dynamics models is massive and two-fold. First, it is a justification after many criticisms from the mathematical (empiricist/reductionist) “pure view” of scientific models about the criteria used in the field (Forrester et al., 1974; Forrester and Senge, 1980; Nordhaus, 1973; Radzicki, 1990; Zellner, 1980) far from being exhaustive. Second, a pure statistical validation process, if restricted to a mathematical formal test, is far from delivering the desired comprehensiveness of a quality enhancement process toward a useful (in terms of the presumed utility) model. Therefore, if one considers an empiricist/reductionist perspective, validation should be done strictly via a formal mathematical process, to which the result would be Boolean (true or false). Validity then, becomes a matter of formal accuracy, rather than practical use (Barlas and Carpenter, 1990). On the other hand, when a relativist/holist perspective is adopted, which is the most appropriate to the present case, the validation of the model becomes something dialogical, iterative, and a process towards learning and participation. In this perspective, models are not necessarily true or false, but open to the new axis of usefulness, under the limitations of a partial, provisional, and socially accepted validity. Here, no model can claim absolute objectivity since every model carries in it the modeler´s worldview (Barlas and Carpenter, 1990).

A historical perspective of system dynamics models validation

This subsection will show the historical review of the recommendations for validation, from the first authors in the system dynamics, to more recent approaches. It will reveal the sources and the rationality for the criteria (Table 1) present in the current validation protocol.

From the seminal book Forrester (1961) produced, three main criteria can be extracted regarding the validity of a system dynamics model: system boundaries, interacting variables, and values of parameters. System boundaries are the most important criteria. Choosing a small boundary creates a system without the endogeneity necessary to understand loops. An oversized system might be distracting and lead to confusion and abandonment of the model. Interacting variables refer to the question if the model embraces the relevant variables and if they are adequately connected. Here, the challenge is to understand if the list of variables used in the model is relevant to the system simulation, but mostly to understand how these variables interact with each other. In both cases, experience is the best guide (Forrester, 1961), which reinforces the idea of modelling being a process of learning and experimentation. Values of parameters are the least important aspect contributing to the validity of a model. System dynamics models are usually lowly sensitive to value variation, and these constant values are only statistically tested after broader and deeper assumptions were already used in the model, such as: objectives were decided, boundaries were determined, relevant variables were chosen, a hypothesis of how each variable interact was created, and an arbitrary level of confidence was established for the statistical test.

The work of Forrester and Senge (1980) is still a main reference regarding a structure for testing the validity of models in the system dynamics field. It is based on three categories of tests:

- test of model structures - before testing the behaviour and outputs of the model, this step of testing will focus on the structure of the model and the system it is simulating;
- test of model behaviour - these tests are focused on the outputs of the model, usually numeric, that can be forecasted or backcasted for a specific attribute of the system. Tests of model behaviour should be applied when the model at stake can produce any type of forecasting or backcasting that allows testing; and
- tests of Policy Implications - These kinds of tests are a comparison of the system behaviour after the application of policies tested in the model, with the outputs forecasted in the simulation.

Meadows (1980) brings a deeper understanding to the discussion, highlighting the relevance of a priori assumptions each modeler translates to the model, conscient or not. The author claims an analyst starts to trust a model when it meets the following conditions:

- Every element and relationship in the model have identifiable real-world meaning and is consistent with whatever measurements or observations are available.
- When the model is used to simulate historical periods, every variable exhibits the qualitative, and roughly quantitative, behaviour that was observed in the real system.
- When the model is simulated under extreme conditions, the model system´s operation is reasonable.

The validation tests according to Barlas (1989, 1996), can be categorized into two main domains: Structural and Behavioural tests. Despite being highly cited in the field, he is mostly concerned with the validation of numerical models and therefore escapes the limits of the present piece. Nonetheless, in the middle of the 1990s, Barlas understood that there was a likely minor complexity type of models, which he refers to as models-for-learning (Morecroft and Sterman, 1994), in contrast to models to “improve performance”. These learning models fit into his understanding of system dynamics in the same way as theory-testing models do, meaning a group of simulations with less rigor, broader participation, focus on learning and experimentation, and finally, with a lower necessity of thoroughly testing: “the models built for learning may not necessitate such behaviour accuracy testing as the traditional applications do” (Barlas, 1996).

The contribution of Balci (1994) comes manyfold. The author conceptualized the whole modelling process in a timely life-cycle perspective. In addition, defined the main types of errors the modelling process can produce, such as: type I – rejecting the model due to lack of quality when it has enough quality to be accepted; type II – accepting the model credibility when in fact it is not sufficient credible; and type III – solving the wrong problem (which corroborates the relevance of scope and boundaries definition beforehand or problem formulation in Balci´s terms). Balci´s test set is divided into five categories (informal, static, dynamic, symbolic, constraint, and formal), disposed of in a crescent level of formality, embracing 45 individual tests. As a good part of these tests are for numerical models, they are not embraced in our case which is focused in qualitative models. Therefore, from this set of indicators, we selected the contributions that are timely, and which would be relevant to the present study (Table 1). In addition, the whole set of indicators will be present as an appendix in the publication derived from this approach. We chose to omit it here to reduce the complexity of an already heavy in content section.

Several of these tests overlap or with minor variations, with each other and with those provided previously (e.g., Forrester and Senge, 1980). Some of the tests, for example, the “review test” is so broadly defined that embraces a whole set of modelling tests. Nevertheless, one of its recommendations regards the timely idea of documentation, and therefore a new line in the test matrix (for the sake of brevity this test matrix was omitted from the present report. It will be presented as an appendix to the paper derived from this report) was created with this specific test.

Another remarkable contribution to system dynamics modelling and validation is the Folding Star Framework (Lane, 1995). In his schematic, the author provides guidance on the system dynamics modelling process based on the following steps: first, an Appreciation of the Situation (AoS) which can be understood as part of the question formulation and some considerations about the status of the system. This part is based on the works in soft systems (e.g., Checkland, 1989; Checkland and Poulter, 2020) with the division of the real system into natural, designed physical, and designed abstract systems (from which cultural attributes emerge and can be considered in the analysis). Second, from the AoS, a Communicated Conceptual Model (CCM) is produced, representing a qualitative representation of the situation, with the clear objectives of sharing the views and problems about the situation with other participants, comparing these multiple views, to finally building an understanding about the system. This step can be used to underpin a mathematical model, the Formal Model (FM). Although this step is recommended by the author, it is not mandatory and the CCM is understood as having enough legitimacy and utility to underpin the final objective of the exercise; third, the creation of the Policy Insights or Recommendations (PIoR), which brings the results produced during the exercise in terms of recommendations to change the system, closing the qualitative loop of the folding star.

This qualitative loop of the star shows three levels for validation: conceptual, inferential, and operational validity. Conceptual validity refers to the coherence of the CCM regarding the AoS, where the ideas of the community about the system must be seen in the model, including cultural aspects (an appreciation of values and ideas the group believes are worth pursuing). Inferential validity measures the extent to which the PIoR can be deducted from the CCM. This is the most fragile part of the scheme and has been a target for criticism due to the big leap required from a CCM to PIoR (in the absence of an FM). Finally, the operational validity is related to the influence the model has in the AoS, forming the feedback of the process into the understanding of the situation. This operational step is more developed for the FM, but regarding the CCM the main ideas concern the realism of the model, the analytical quality of the PIoR, and the satisfaction felt with the process. In general intermediate-quality states are considered reasonable targets due to the lower analytical potential of qualitative models (Lane, 1995).

Some remarkable aspects of the folding star to the present piece are, first, to bring to the discussion a group of variables related to cultural assumptions in the model (naturally qualitative). The cultural assumptions tests concern the social elements of a system dynamics activity. These elements of investigation try to embrace the "different perceptions of a problem that might exist in and to address the social realities of the group" (Lane, 1995). To the author, exposing the differences of opinion regarding the problem is crucial before converging into the problem statement. As the AoS and CCM are plurally discussed, a deeper appreciation of values and other cultural aspects embedded in the model is expected, possibly resulting in meaningful recommendations to PIoR. Second, it brings a set of tests regarding the usefulness of the modelling process, including but not limited to, the costs and time involved, the social/political capacity to implement the recommendations, the affinity between the participants with the modeler and with the modelling process, amongst others. This meta-understanding of the modelling process might bring relevant information regarding the satisfaction of those involved in the modelling exercise and crucial aspects related to the social practice enabling or obliterating the implementation of the recommendations discovered by the simulation process.
These ideas of broadening the participation in problem formulation had their roots probably with the foundation of operational research with Churchman et al. (1957) (Reisman and Oral, 2005), but during the 1970-1980s, they were markedly reinforced and formed the emerging field of the soft systems methodologies (SSM). This school of thought is concerned with the plurality of social participation in the modelling, as stated: "What is the system? What are its objectives? ignore the fact that there will be a multiplicity of views on both, with alternative interpretations fighting it out on the basis not only of logic but also of power, politics, and personality." (Checkland, 1989).

The main point defended by the author is that Systems Analysis, Systems Engineering, and Operational Research, despite small variations, deal with the same thing, namely well-defined problems. To these problems, the elegant solution (i.e., optimal or efficient), suits the goals of the modelling process. On the other hand, SSM goes for a messy, ill-defined, frequently contested terrain, where the elegant solution rarely will be the answer since conflicting worldviews act upon the definition of the problem and the solution. This broader understanding is congruent with more modern views of system dynamics, considering there are no value-free theories and no value-free models (Sterman, 2000).

Another important trait in SSM that makes this perspective suitable for social-ecological systems modelling is that the problem is never taken as permanently solved. The solutions obtained, by a process of accommodation of those conflicting worldviews, are always provisional since the group of assumptions that were considered in that negotiation can change, namely, the state of the system, the balance of power, the emergence of new problems, or even the worldviews (Checkland and Poulter, 2020). The main objectives of SSM then are to organize the process of discussion towards the solution of a problem (purposeful action), in a constantly learning perspective, that produces solutions both desirable (in terms of the options given/structured by the decision-making process) and feasible (meaning tolerable, considering the conflicting worldviews). From this perspective of SSM, more recent branches of systems thinking emerged, such as: holistic flexibility (Chowdhury et al., 2023).

The quality assurance process of a SSM, is very basic, embracing the ideas of: efficacy – whether the transformation in the system is producing its intended outcome; efficiency – whether the transformation is being achieved by using the minimum amount of resources; and effectiveness – if the transformation is helping to achieve a long-term or higher-level aim (Checkland, 1989). Other elements can be added to that, such as elegance, understood as an aesthetic criterion, differing from the optimal criteria as in system dynamics; or ethicality, which questions the ethical foundations of the transformation proposed (Checkland and Poulter, 2020).

In a broad sense, the author recommends users check the model for coherence, and uses the term defensible, instead of correct, to name a model that passed this coherence test, to which extent every connection in the model is meaningful in terms of understanding the Root Definitions (boundaries) and the CATWOE (a mnemonic for the process of modelling, meaning C: customers, A: actors, T: transformation, W: worldviews, O: owner, E: Environment) (Checkland and Poulter, 2020; Haynes, 1995). These three elements of quality assurance (efficacy, efficiency, and effectiveness) were understood as being congruent to the system improvement, time-and-cost of investigation, and family member tests respectively, to which the worldviews/values and boundary-related tests were added considering the central ideas in SSM.

To the limits of our knowledge, the issue of validation/verification of system dynamics models got cold during the last 20 years, with much fewer publications. Exceptions made by some articles (e.g., Lemke and Łatuszyńska, 2013; Qudrat-Ullah, 2012; Qudrat-Ullah and Seong, 2010) that reproduced Forrester´s 1980s ideas with the improvement of the mathematical approach for quantitative tests, but with minor theoretical increment. The work of Andersen and collaborators (Andersen et al., 2012) called attention due to the focus on validating qualitative models. In their contribution, disconfirmation (i.e., invalidation) of a causal construct can be made by asking a third party, not involved in the modelling section, to create a judgment about the model, where the level of agreement with the simulation corresponds to the level of validity of the model. We argue this approach can only make sense if the worldviews and values of both people interviewed during the model building and disconfirmation stage are congruent. If we assume building causal models are similar to a theory creation (Forrester, 1961; Sterman, 2000), and worldviews are determinants of the way people frame their understanding of the world and its problems (Ney, 2012; Thompson et al., 1990; Verweij et al., 2006), it would be expected that people with conflicting worldviews would disconfirm previous models not based on their validity to simulate a problem, but due to the differences in how they frame problems. That was already recognized by part of the system dynamics community (Checkland, 1989; Checkland and Poulter, 2020).

The most recent review (Schwaninger and Groesser, 2020) brings some important elements of validation, including a useful loop-dominance idea, but also ignores SSM and the debate about embedding multiple rationalities in modelling. It is also more dedicated to the validation of quantitative simulations, following the traditional approach (e.g., Forrester and Senge, 1980).

Notes about semantics

The discussions regarding the quality of system dynamics models usually run around the terms validation and verification, and according to Lane (1995) started with the seminal work of Ackoff (1956). The foundational book of system dynamics, Industrial Dynamics (Forrester, 1961), has a full chapter named “testing model validity”, showing that validity and verification, and most importantly, how to achieve them, are central to the discussions in the field. In the present subsection, we will discuss what these terms means, the differences in validation and verification, their common uses and limits, and why we adopted the name “presumed utility” for the current piece.

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 modelling 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 modelling 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.

To Forrester and Senge (1980) validation is the process of “establishing confidence in the soundness and usefulness of a model as a policy tool”. To that end, the confidence in the model must be transferred from the modeler to the users, a step without which, the model's potential to enhance the management system will not be realized, and thus the model is useless (and invalid). The authors also suggest that there is no ultimate proof that a model is correct. What can be done is to prove it is incorrect, and therefore by surviving many tests, results in a reliable model (which is different from an absolute truth), a perspective that seems aligned with Popper´s falsifiability principle.

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.

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. This practice is not possible with CLDs, since these are qualitative simulations of the system.

A common quote in modelling institutes is that “all models are wrong” (Sterman, 2002), usually complemented by “but some might be useful” (W. Edwards Deming in Van den Belt, 2004). 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 towards the notion of “usefulness”.

Forrester (1961) arguments that “the final test of utility happens when the model unfolds into a better management system” still applies, but as it can only be determined with time, the present evaluation protocol adopts a terminology to name an incomplete task, as it is dedicated to evaluating something that is not fully deployed, still provisional. Echoing Meadows (1980) that the term “utility” suits better for an indicator of the quality of a model, we adopt the term “presumed utility” as the indicator of the quality of CLDs. 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.

Appendix 2: Follow-up questions of Model and stakeholders information

A: What is the main message the model is passing?

(Macaronesia)

1. The model shows the intricacies and inter-dependency of model parameters highlighting the complexity of social-ecological systems.
2. Implementation of protection measures, both in the ecological corridor Azores-Madeira-Canaries and in coastal MPAs, impacts local economies (i.e. tourism) and vice-versa.
3. MPA conservation; Ecological corridors between archipelagos
4. The model allows to explain in a simple way the complexity of the system. Interconnectivity among the variables. Highlights the need of conservation and protection measures.
5. That there are specificities and constraints related to the marine environment where the DA lays because it is mainly open ocean, thus more sensitive to certain aspects such as boat traffic, deep-sea mining, and more robust to others such as climate change effects.
6. The model passes a message of the system functioning in the Macaronesia, revolving specially around the nature of tourism activities and the possibilities of an ecological corridor that focused on the migration of megafauna. PESTLE approach brought up the main issues and opportunities of different sectors for the achievability of a sustainable scenario.
7. The potential for the implementation of protection measures in the DA, and the possibility of connecting the Macaronesia area through a corridor. I also believe that the model represents well the causality between social, economic and environmental factors in the area.
8. The main message that the model is passing concerns the different pressures of tourism in the DA and the biological corridor that connects all three archipelagos involved.
9. CLDs are a powerful tool for understanding complex systems and identifying points of influence for change and making-decisions. This model helps to visualize how different variables influence each other and create feedback loops that reinforce or balance system behavior. In a simple way, the model provides valuable information for decision-making in improving the balance between the conservation of marine biodiversity and the economic activities developed in the marine environment (special role of ecotourism).
10. It emphasizes the ecosystem's complexity and the necessity of implementing conservation and protection measures in the ecological corridor and MPAs to achieve healthy habitats.

(Arctic)

11. How governmental decisions and actions impact both nature and socioeconomic systems.
12. It is difficult to see the main message from looking at the model. One reason could be that the model is not complete. For instance, loops are not communicated (labeled) in the model

(Tuscany)

13. Tourism emerges as a fundamental element for the Tuscan Archipelago community that has positive effect in terms of job opportunities but may also have negative effects on marine ecosystems (and the services they provide) and wellbeing of island residents.
14. This model is suggesting the pros and cons of tourism in the Tuscan Archipelago DA. Specifically how tourism affects key ecosystems and social systems and how they interact and feedback on one another.
15. Tourism activity is the most important economic activity for local people in the different islands, but it could also have negative impacts on marine communities without clear roles. Socio-economical activities and the environmental conservation should be strictly linked

B: What policy insights or recommendations can be created meaningfully?

(Macaronesia)

16. One of the main concerns raised is that an ecological corridor among archipelagos would cross many geopolitical boundaries (e.g. regional, national, international) and that this needs clear articulation with the different agents, which may not be easy to accomplish.
17. We did not enter into specific details. We discussed about implementation of protection measures in the ecological corridor Azores-Madeira-Canaries. From the CLD we can suggest that such measure can be adopted and maybe can be accepted more likely if technological development is "moving" in the same direction. For example, development of less noisy engine, or detection system to avoid collision with cetaceans.
18. Policy insights, as many as the stakeholder's. Not sure about recommendations.
19. Recommendation for heathy ecosystems and habitat quality. Protective measures. Sustainable tourism.
20. Supranational ones concerning a possible marine corridor and its management.
21. Some insight can be created, such as the rearrangement of marine traffic to prevent environmental impacts, the implementation of innovative technologies for increased sustainability and potential cost reduction, as well as the evaluation of size restrictions and fishing campaigns for fisheries, based on scientific evidence of size to sex ratio and reproductive activity of some species. Additional insights include shifts in tourism

activities to a more sustainable and local economy based approach. Despite the potential of these considerations, further evidence and resources might be needed to back them up.

22. A recommendation for the establishment of an ecological corridor for Macaronesia, and maybe to propose some transboundary MPAs in the area.
23. The increase in regulation regarding overtourism, and the creation of an MPA covering the biological corridor.
24. The possibility of creating a ecological corridor at a supranational level has the potential to improve policies associated with the management of small island systems and their particularities.
25. The recommendation keywords are regulation, protection, and conservation. Regulate and protect MPAs and the ecological marine corridor considering human pressures, to ensure healthy and sustainable ecosystems

(Arctic)

26. I have never participated in any work related to policy briefings or recommendations and I have no idea how to turn the model into a policy brief that makes sense to the intended audience.
27. There are no new insights that come from the model, instead, it confirms current knowledge.

(Tuscany)

28. Regulation of touristic activities within the MPA but also outside. Deployment of buoys in the most visited places/diving sites for private boats. Dissemination and educational activities to inform the public audience about best practices at sea, marine environment and species in the Tuscan Archipelago. Public funding to help locals to shift from extractive activities (fishing) to more sustainable tourism-related jobs.
29. One recommendation which arose from the modelling process was creating a series of apps to engage with the public in order to let them make informed decisions about their tourism activities in the archipelago. Furthermore, instituting policies aimed at regulating the number of tourists was mentioned.
30. Local people and tourists should be adequately inform about the importance to conserve marine biodiversity and key habitat forming species. More information about this topic are needed.

C: How much novelty was learned or discussed during the process?

(Macaronesia)

31. I felt that most of the exercise was fairly intuitive, but it helped to have a structured and well thought-out step-by-step process for the model creation.

32. The novelty was the use of CLD model.
33. All the discussion process provided more information and novelty.
34. New way to approach a complex system
35. In the end I believe we had a deeper understanding on the strengths and weaknesses, challenges and opportunities, of the DA as an all.
36. Some elements of novelty were brought up during the modeling process, which included recent technological advancements, as well as proposal for the creation of MPAs and scientific evidence for the creation of a Macaronesian ecological corridor, along with some of its limitations. Project opportunities for ecology studies were also discussed.
37. The CLD and PESTLE models were new to me.
38. The three archipelagos had the opportunity to learn about the realities of the other locations.
39. In addition to the list of variables for the design of the CLD and the relationships between them, the PESTLE analysis provided a lot of valuable detailed information on the context of each archipelago and discussing common points of the Macaronesian DA.
40. The new thing was the interactive creation of a complete model.

(Arctic)

41. Leaning how to qualitatively evaluate impact of fishing on the socioeconomic system and how governmental decisions impact many things.
42. Due to time constraints and the simplicity of the model, very limited novelty was discussed or learned.

(Tuscany)

43. The process can be very useful to start a constructive discussion among stakeholders/modelers with different expertise/ point of views. It was also useful to think about one issue under different perspectives (ecological/economic/social) at the same time.
44. I for one saw the whole process as novel, I am unsure as to whether these topics have been discussed before
45. We are often focused only on research activities and sometimes we do not invest too much time to divulgate the results of our activities outside our "boundary". This process highlights how science and tourism (or economic activities in general) ca be strongly linked.

D: Do you think the knowledge represented in the model is potentially acceptable for your Stakeholders?

(Macaronesia)

46. For most stakeholders, yes.

47. The model is a good tool to explain the system. I think stakeholders will easily understand how everything is connected and related. I'm not sure if this will make stakeholders more supportive of new management actions.
48. Yes
49. Probably, but it should be presented and explained in a simple way according to the stakeholders group
50. To some of them yes, but to others maybe not; it depends on their individual wisdom.
51. The knowledge presented is potentially acceptable, although this may vary depending on the type of stakeholders involved. This might be the result of the modeling team being mainly comprised of biologist. In any case, adjustments could be necessary once stakeholder feedback is received.
52. Yes, I believe it will be acceptable for most stakeholders. However, I still feel that we might have missed the inputs from someone outside of academia during to exercise, to strengthen the model.
53. Yes, if well explained and contextualized.
54. I think so but, in addition to their perspectives on the problems and challenges that arise, it will depend on how the storytelling of the CLD is carried out and how the information is transmitted in an accessible and connected way.
55. Yes, I think the model has potential, but it should be explained in a simplistic way and explanations developed according to stakeholders' specificities and levels of knowledge.

(Arctic)

56. I have not been involved in the stakeholder interviews therefore I have no idea what is acceptable to them and what is not acceptable.
57. The model needs further refinement, such as a clearer illustration of loops, to be used for communication purposes.

(Tuscany)

58. I think the model can be generally acceptable, some stakeholders may disagree with individual elements (or links) or may think the model is incomplete.
59. I think the knowledge would be acceptable if presented in a palatable and easily digested format.
60. Yes, I think so, but maybe one single CLD contains too much information, and it is not easy to interpret it.

E: How do you evaluate the relevance of the information of D2.1 in your knowledge base used in this model exercise?

(Macaronesia)

61. I learned quite a lot since I started working in the project (not only from the D2.1) since my background as a field ecologist is quite different.
62. I have read the D2.1 but to be honest I do not remember it. Since I have actively participated in the WP2 interview/survey it is possible that the knowledge acquired during those exercises has influenced my awareness of the system. So, I believe that having previous experience might have facilitated the construction of the model.
63. It was important, but not capital.
64. It was relevant somehow since it was based in the same approach as in the WP2 stakeholders' interviews, although our model was developed in a different perspective and considering a specific topic.
65. Some of the elements presented by the stakeholders during the interviews, were shared in our model too. Still, their answers did not influence my line of thought.
66. In accordance with the impressions left by the Canary Islands stakeholders interviewed for D2.1, the importance of cultural heritage linked with socio-economic activity was brought up during the modeling process. Mass tourism was also a relevant topic, along with its interaction with other variables, particularly those of ecological value. The autonomous status of the Canary Islands and its impact on governance was likewise brought up in discussion, and although climate change was not included in the Casual Loop Diagram developed, it was mentioned as a limiting factor for the achievability of a sustainable socio-ecological scenario (SSP1-RCP2.6), with how it could interfere with the creation efforts of a Macaronesian ecological corridor.
67. I read only the summary of D2.1, so I don't feel like it was relevant for my knowledge to use in this model exercise.
68. It was relevant because it gave me a bigger perspective (being able to see the "big picture" of the DA).
69. In my case, I joined the project recently. I assume that the relevance has also been addressed by including feedback from the stakeholder workshops and is reflected in the relationship between variables in both, CLD and PESTLE analysis.
70. The information in D2.1 probably contributed to recalling knowledge and issues, helping to build reasoning

(Arctic)

71. It was irrelevant as I did not participate in task D2.1 and I did not read information from D2.1 before the workshop
72. Whilst the information from the Arctic DA in the Annex of D2.1, is informative, the method of building the model does not utilize all information from D2.1.

(Tuscany)

73. D 2.1 was useful to better understand differences and commonalities among stakeholder views.
74. I think the information contain in D2.1 is highly relevant and provided a background to be used within this exercise
75. It was useful to deepen our understanding about local communities and economic activities in the TA

F: Did you use direct information from the report on stakeholders (D2.1) to frame your understanding of the system represented by this CLD?

(Macaronesia)

76. Yes. A substantial part of what was made here results directly from what was learned when contributing to D2.1.
77. Direct information no. Indirectly maybe yes (See previous answer).
78. No.
79. No, we did not consider the stakeholders inputs, however this model should somehow interconnect/or validate with stakeholders' opinions
80. No, I didn't. As I said my mind was already made up and if we share some thoughts, that means we share the same concerns.
81. The existence of an agreement in relevant topics between the stakeholders and Canary Islands representatives for the workshop allowed for the information of the stakeholders report to be applied during the modeling process.
82. No.
83. I think not in a direct way, but expert opinions already have interactions with local stakeholders, and that is partially reflected in the results.
84. No, I didn't. The previous documents read to contribute to the development of the model were those previously provided by e-mail by AZTI, to understand the different elements of the CLD and its practical application.
85. No, maybe indirectly.

(Arctic)

86. No, I did not.
87. Information from the report on stakeholders (Arctic DA) is to some extent represented in the CLD, but there is potential to further analyse the results.

(Tuscany)

88. Yes
89. I did indeed.

90. In general, I know the social-economical activates of different islands of the TA and so I was able to work on it also without reading the report on stakeholders.

G: How do you think the outputs of WP2 can be integrated into the product of this workshop? (Macaronesia)

91. I believe that the model should be presented and validated by different stakeholders to ensure its accuracy and holistic nature.
92. I think the two approaches can be compared to assess if the view from the stakeholders (data from the WP2) is similar to the one of the researchers (CLD model). However, two important notes: a) In our DA researchers who build the CLD were also interviewed or took part in the WP2 work. So, results from WP2 may have influenced the construction of the CLD. B). In our DA, WP2 never focus the interviews on the ecological corridor. So, this aspect was never discussed with our stakeholders.
93. One way to incorporate the outputs from WP2 could be taking into account the scores (weighting coefficients) in the interconnections, but there's a need to considers the diverse stakeholders' groups.
94. By searching for discrepancies and similarities, between the stakeholders' opinions and our own.
95. WP2 outputs could perhaps be integrated into the product of the workshop by contrasting stakeholder interviews with the discussions held by the modeling team during the workshop, searching for relevant topics, links and implementation possibilities. Also, harmonization of the elements brought up according to deliverable 2.1. Important contributions could also be made by presenting the CLD built to stakeholder groups for feedback, tuning and refinement. The PESTLE analysis could also be enhanced in the same way, at the very least allowing for the reception of feedback.
96. It can be helpful to compare both and identify the similarities and differences and gaps between the two approaches.
97. Results from WP2 gave us the perspective of the stakeholders' main concerns which is reflected in the model, although not in a very detailed way.
98. With a participatory practices approach, the implementation of the CLD would provide valuable information for decision-making that, in defining recommendations and improving policies, would facilitate the harmonization and integration of simple socio-ecological frameworks. In this sense, it is absolutely necessary to apply a dialogical process that allows analyzing the level of adaptation of the model designed to the needs and challenges identified by stakeholders.

(Arctic)

99. I do not know.
100. You will have to directly analyze the results to complete the CLD model. However, the qualitative information from D2.1 is not so compatible with system/CLD, so methods of handling information from D2.1, need to be further explored.

(Tuscany)

101. It is possible to make comparisons between CLD realized during the workshop and those created by stakeholders.
102. I think maybe as a background or introduction framing the different DAs.
103. The outputs of WP2 could deepen our understanding about the request of local people that live and work on the islands and their points of view. As I write above, I think it is really important to integrate our knowledge with the necessity of local people and tourists.

Appendix 3[2]: Comments from the Demonstration Areas (DAs) after the workshops for the Causal Loop Diagrams (CLD) development

Type	N	Criteria	Comments
Guidelines and processes	1	Purpose	1. The purpose of the model is clear 2. Yes 3. Not perfectly clear (grade 4) 4. Yes, It was clearly presented. 5. It still seems a little ambiguous in certain aspects 6. The purpose of the model was explained at the beginning of the first session, although perhaps some more time could have been dedicated to this, even prior to the workshop. 7. Yes 8. The purpose of the model was very clearly explained. However, in order to facilitate a more in-depth analysis of the aspects linked to the exercise, it would have been of great support to receive in advance the document explaining the context and detailed content of the components of the proposal, both CLD and PESTLE. 9. I understood that the model was being built as it was needed for the work progress of another work package. 10. It remains unclear how the model will be utilized to stakeholders or how it will inform future tools. 11. Although our DAs working group is comprised of ecologists, the rationale behind the model building was clear from the intial explanation. 12. Yes
	2	Usefulness	13. The model is clear to us (modelers). It may need clarification/introduction to third party users 14. Both the modelers and 3rd parties. Not if presented without a full description of each model elements and a brief introduction. If presented with a introduction to the study area and a description for each elements yes to both. 15. Both; You should provide more information; You should provide more information; You should provide more information; 16. Modelers will develop the model, adapting for each site, and 3rd parties will use the model to understand the system. A full description of each model elements and a brief introduction should be provided to a better understanding. With the a briefing/introduction they will understand the model, but for use or make advantage of it will be need a better approach to transpose from the model to the reality. 17. The model should be for third parties. The format maybe easily understood by most end users, but not all thus, not all end users will be able to use the results of the model. For instance, the background of the user is of outmost importance to understand and make use of the model.

[2] These comments were not edited for clarity or grammar edition. They are exactly the words the DAs sent as feedback, including the orthography mistakes.

			18. The model was constructed with its usage in mind, for both modelers and third parties, and adjustments were made when the practicality of the model was brought into question. Despite this, introductory explanations might be necessary in some cases. 19. I believe it can be used by both, the modelers and third parties. I'm not sure if all third parties will be able to understand the model and use its results. 20. The model is useful for the modelers, but it has a biased perspective from the experts. 21. I believe that the model is well suited to users and to decision-makers, in that it includes qualitative information from the storytelling of the designed diagram. Furthermore, the PESTLE analysis carried out allows an approximation to those factors of the global context that will affect the projection of the model, and its effective implementation on a local scale. 22. I have no idea who will operate the model, hence I have idea if they will be able to understand and use the model and its results. 23. Considering the purpose being rather nuclear, it is difficult to assess how operators or future model users will interact/apply with the model. The model is the results of a two-day workshop. There therefore remain obvious limitations in its capabilitites. 24. It wasn't fully clear to me who should be the final users, but I think the process of building the model can be very useful. 25. The application of the built model to third parties would require additional simplification as the model is quite complex and information would be difficult to access 26. It is in an adequate and clear format for most users, but it still needs to be explained, adapted for some stakeholders, and tested with all stakeholders. Before being presented, an action plan on how to present the model according to stakeholders' capacities should be developed.
	3	Presentation	27. The model is fairly simple to read and should be intuitive to most people 28. No maps or figures were used. Several loops are represented in a simplified way. If the model is presented with a introduction to the study area and a description for each elements, is easy to understand. BUT we did not try to present the model to 3rd parties. 29. Maybe you should present easy understandable maps of the areas. This needs to be tested outside the group to identify snags. 30. It depends on the audience (it was not present to any audience so far) and requires a good explanation about the model proposed. Individual loops should be presented when addressing to a specific topic. If its need to understand all the system there is a need to present in one big CLD. Several loops are represented in a simplified way. Some crossed lines can be double checked in order to reduce more is possible. 31. If by audience you mean us, the answer is yes. The loops, although individually presented, in the end are difficult to individualize in the CLD. The number of cross lines were reduced as much as possible but are still too many. 32. The model was constructed for its intended audience, regarding links between variables and the names chosen, and it was simplified when deemed possible. However, diagram organization was left to be improved, and no maps or figures were utilized. Loops were represented in a big general diagram. 33. If we are considering the relevant audience, then yes. For third parties, I think we will need to add more information in the introduction, regarding the DA and to use more infographics to be more appealing.

			34. The diagram is not simplified, and it can be hard to follow. 35. The model was presented in an adequate manner, supported by a ppt presentation and clearly explaining the key aspects for the development of the exercise. Some examples of simple diagrams were projected to understand the relationships between variables and to try to respond to the causality vs. correlation approach. As for the presentation of the loops, they were represented in a single CLD, which may be somewhat complex when developing the storytelling of the relationships between variables. Being a dynamic tool, it might be appropriate to analyse the simplicity of some of the information included in the CLD, in order to avoid redundancies in the variables and their content. 36. I only have the Vensim model diagram at hand. The model has several loops. It a busy web of circular connections. The language used is mostly general language with one specialized scientific term (Maximum Sustainable Yield). There appears to be four crossed lines. 37. The adequacy of the model is presentable, using little scientific language. However, lack of materials 38. The model is represented as one large CLD. As mentioned in the previous comment it would need to be greatly simplified for the audience 39. The model was constructed based mostly on team members' ideas and knowledge, and it has not yet been presented to stakeholders. Not all stakeholders will understand it.
	4	Perspectives in Boundary-adequacy	40. The model supports the views of the people that were present, which were fairly discussed, but it may be biased towards their view since the people involved in model creation only included few different stakeholders. 41. No b) In a simplify way c) Yes d) Potentially yes, if each policy is an element of the model 42. No; No; Yes; Not really 43. Only on the scenarios presented. Although it does not take into account the stakeholders' perspectives. 44. Scenarios presentation b) Yes, but without taking into account the strength of each variable. C) Yes they were presented d) No 45. The different perspectives were actually biased in our case, since we were all academic and mostly biologists. Qualitative models are always a bit speculative so, I believe this model encompasses the most important aspects to be addressed with this type of research, and allowed discussion on the subjects mentioned here. 46. The modeling process supported healthy discussion throughout, reaching agreements and promoting the participation of everyone involved. A) No, B) Yes, C) Yes, D) Yes. 47. Yes, the model supports debate from different perspectives. However, in our case, I believe we here missing some relevant stakeholders to have a more holistic perspective on our model. Most of us are from academia, so we were a bit bias in the way we defined the CLD. 48. Although several experts discussed the approach, most of them were biologist working in conservation/restoration, which can biased the final result. 49. This adaptation has been possible, both in the discussion during the design of the CLD and during the analysis carried out from the political dimension (and other), within the PESTLE. 50. As a biologist, I am lost in the questions posed here. 51. It would be useful when building these models to have a mixed group of backgrounds to provide counter arguments to specific groups. As our DA is comprised of ecologist there is little disagreement in the model building.

			52. The model can support debate. no; b) yes; c) yes; d) very briefly
	5	Norms/values in boundary adequacy	53. The model supports the views of the people that were present, which were fairly discussed, but it may be biased towards their view since the people involved in model creation only included few different stakeholders. 54. I think this is NA because we did not test the model with 3rd parties. For us (the modelers):a) yes; b) yes so is a 5 55. Yes; Yes, but needs to be tested outside the group 56. It does not take into account all actors /stakeholders sector perspective. Biased model 57. There is no doubt that the model allows for the debate of different perspectives and values. Our opinions are tailored by our culture so this will always influence our vision of a desired state. 58. As stated for the previous indicator, debates for this topic were held in an organized and polite way during the modeling process. However, no third parties were involved. A) Yes, for modelers. B) Yes, for modelers. 59. I assume that this aspect has also been addressed by including relevant feedback from the stakeholder workshops, and is reflected in the relationship between variables in the diagram. A detail to highlight is the importance of transferring the content for consultation with all stakeholders, so that the suitability and applicability of the model in different contexts and situations can be determined from a dialogic approach. 60. As a biologist, I am lost in the questions posed here. 61. Certainly, it will support debate, but the debate with stakeholders has not happened yet. The model was built by biologists, and some biases need to be assumed. However, incorporating stakeholder opinions/discussions can suppress this bias and contribute to validating the model.
	6	Trustworthiness or Guru status of the system dynamicist	62. I think this is NA because we did not test the model with 3rd parties. 63. I think it will depend up on the stakeholders and modeling team actions. 64. Yes only with the perspective of the practical exercise made with the team. But if this question is related to stakeholders, the answer should be "0 – not apply". 65. I did not now the modeler prior to this exercise, but he was able to develop a positive atmosphere among the group 66. Stakeholders were not involved in the modeling process, thus positive feedback could not be reported. 67. We didn't test this model with stakeholders. 68. It's possible to establish a good relationship between stakeholders and the modeling team, but it didn't happen in this exercise. 69. The assessment of this indicator is based on the good dynamisation carried out by the person in charge during the workshop. When talking about stakeholders, we consider the participants belonging to partner entities of the project, which, from the relations of the consortium in the DA of Macaronesia, have promoted a continuous and fluid dialogue, based on a code of conduct established at the beginning of the first session. 70. Overall experience of working with pelagic fishing industry stakeholders in my home country is positive relationship. This applies to several projects outside scope of Marine Sabres 71. Hard to say, because I am not in direct contact with the stakeholders, plus we did not have the possibility to show those result to anyone yet 72. The model was not yet presented to stakeholders

	7	Meaningfulness of the process	73. It was a fun exercise to create the modeler. So I can only reply as such. 74. I think this is NA because we did not test the model with 3rd parties. But I had a lot of fun building the model! 75. Yes, not much; not much 76. 0 – not apply, since it was not presented to external stakeholders. And 4, if the team counts as stakeholder in this model results. 77. There were no stakeholders in the modelling exercise 78. Stakeholders/actors were not involved in the modeling process, so they could not participate in discussion. 79. We didn't test this model with stakeholders. However, I enjoyed the workshop and exercise we made together. (If this WK will continue to be online, my only recommendation would be to use tools such as Canva, Miro, Padlet or others, since they allow participants to add information and comments at the same time as the discussion runs. This allows the organizer of the WK to receive feedback from everyone, bearing in mind that in an online call, it is sometimes difficult to have the opportunity to speak without running over others.) 80. Only the modeling team participated in the process. 81. It was really a very enriching process, with great participation of all attendees, and with a debate that incorporated aspects from different knowledge and sectoral approaches, providing specific examples of each of the archipelagos that make up the Macaronesia DA, within the framework of the project. 82. Stakeholders did not participate in building the model. 83. Hard to say, because I am not in direct contact with the stakeholders, plus we did not have the possibility to show those result to anyone yet 84. Again, in the current form I believe the model would be difficult for stakeholders to access. I think a simplified CLD or a series of worked scenarios or examples would be useful for stakeholders 85. The same as above [The model was not yet presented to stakeholders]. However, the process of construction of the model with the team was good.
Specific model tests	8	Structure-verification	86. I am happy with the breadth and accuracy of the model, but as mentioned earlier, it may showcase a biased view of the system. 87. A) yes B) I hope so, yes C) yes D) we don't have delays 88. Yes; Yes; Yes; Yes, but I think we are on schedule 89. A) The model represents as much as it could possible the structure of the real system. B) For some variables we can say that stated unambiguously, but for others there is a need to explain better the variable used otherwise it can cause some ambiguously. C) Yes D) No, Delays are not represented 90. I believe the model represents satisfactorily the system and the variables are clearly stated. The connections had + and – signals implying causation. As for the delays I have no opinion 91. The model represents the system and its issues satisfactorily. The variables are generally stated unambiguously, although some could be improved. Connections represent causation, but delays in response were not incorporated into the model.

			92. I think that the model has resulted in a good representation of the system, however, I feel that perhaps some social/economic aspects are missing or misnamed. In other words, they are represented to a certain extent, but perhaps we haven't been able to name them in the right way, so that they can be more easily interpreted by third parties. Causation is well represented. 93. There was some difficulty in determining the inclusion of some variables within others, so I am not sure whether, at the time of storytelling the diagram, there might be some ambiguity in explaining to users of the model, the focus of those variables. An example of this is the concepts of carrying capacity and heritage (tangible and intangible). As for the establishment of 'causation' and non-correlations between variables, from my point of view, the previous instructions provided by the facilitator, and the fluency of the discussion allowed the relationships between variables to be adequately established. With regard to delays, I do not recall that symbolism has been included in this sense, reflecting delays. However, temporality was very much present in the analysis. 94. Yes, I consider the model structure represents the real system and its issues. 95. The model represents a highly simplified view of the System although any more complex and it would be unwieldly/impractical. Without data and experimentation the connections simply represent an assumed correlation. 96. In my point of view connections represent correlation and only potentially they can represent causation. 97. Yes
	9	Loop Polarity	98. I think this is NA because we did not test the model with 3rd parties. 99. Needs to be used outside. 100. We did not test the model with other stakeholders/"clients". 101. There were no stakeholders in the workshop, apart from us, academics 102. With regards to the identification of relevant loops by stakeholders, they were not involved in this process. Polarity was properly determined and there was convergence into variables of interest. There were clear goals, but loops were not named. 103. We didn't test this model with stakeholders 104. Not all stakeholders identified the relevant loops, just the modeling team. 105. I think this indicator does not apply since the workshop was not addressed to stakeholders but to members of the Macarosia DA consortium. 106. Stakeholders did not participate in the model building. I consider loop polarity properly determined. Loops are not named. 107. Depends on the time scale (long vs. short term) 108. The tool has not been represented to stakeholders? 109. We did not focused on loops during the workshop. 110. The model was not tested.
	10	Boundary adequacy (as structure)	111. There is some level of aggregation in order to keep the model simple. 112. Some variables are aggregated to simplify the model 113. Yes, most of them 114. Some variables 115. There was a good identification of the relevant variables, and although some of them could be overlapped, there was a quite good representation of these relevant variables

			116. The aggregation level of the model constructed was adequate and relevant structures were accounted for. Relevant variables were incorporated explicitly. 117. Since we had to simplify the model, some variables are aggregated with others. 118. The variables are well represented 119. Although it is possible that there may be some overlap between variables, I believe that the relevance of the variables included in the diagram was very well agreed and analyzed 120. I consider all important variables explicitly represented 121. I think the most relevant elements have been included. We also avoided to be too specific and we aggregated some similar elements. 122. To simplify the model some variables were masked. However, this may be a positive for any attempt to engage with stakeholders 123. The model represents well the structure of system
	11	Family-member	124. I believe the model can be applied to similar "problems" or similar geographic areas (e.g. insular regions) with minor adequations. 125. Yes B) Yes 126. A) yes b) yes with some careful 127. Absolutely (grade 5) 128. Model application for minor adequations was not tested. However, the model could retain applicability at a more local scale with minor adjustments. 129. Yes, I fully agree with the description of this indicator. 130. I do not have the knowledge and experience to answer this question. 131. I am not sure, this is a very specific case in the end 132. Yes
	12	Extreme-conditions	133. Not sure I understand this, since the model is not numerical. 134. ? numerical? It is qualitative. Yes, it will behave the same in extreme conditions. Yes, it is possible to infer this 135. I have doubts and lack capacity to come to a conclusion regarding this. 136. Being a qualitative model so far, yes it could be done, but there is a need to adequate the model for extreme-condition if its quantitative (weighting coefficients) 137. Our model was not numerical 138. Model behaviour under extreme conditions of variables was not tested or considered during the modeling process. Nevertheless, given the present model, appropriate functioning could be inferred. 139. Numerical model? 140. I think that under extreme conditions and in a numerical approach, the complexity of the analysis and the deviations in causation between variables, would make inference very difficult. 141. I think so. 142. Not sure about this

			143. This should be tested 144. In general, I think this approach can be hardly used to make inferences. 145. It is possible, but not as it is. The variables/connections between variables should be pondered and scores added to the interactions
Policy insights and spillovers	13	Insight generation capacity	146. The model still needs validation from different stakeholders in order to test its applicability or ability to generate recommendations. 147. Not really because we did not directly recommend policies. BUT is it possible to infer that policies are needed to regulate some elements 148. I do not think the model leads to this 149. We did not directly recommend policies. But it is possible to infer some recommendations and policy needed to regulate some variables. 150. It is hard to do any policy insight or recommendation strictly based on our model 151. Discussions during the modeling process led to some insights and considerations regarding the influence of geopolitics, protection of key species and the Macaronesian corridor, among other topics. 152. I don't believe we discussed policy recommendations yet, but the model sure does have the potential to help build some PIoR for the DA area. 153. It wasn't tested yet, so no. 154. Although not applicable in our case, if the model were to be implemented through a participatory process with a multi-stakeholder approach, I believe that this tool would be very useful in identifying recommendations and influencing policy and decision-making (strengthening governance). 155. No, it did not. 156. Not yet, it just captures the current state so far 157. It is possible to derive some recommendations
	14	Relevance and Fertility of PIoR	158. The model does imply that in order to accomplish its purpose, a few policy chances may be required owing to the geopolitical nature of the system. 159. I do not think it should lead or if it is relevant 160. Policy insight or recommendation were not established 161. Does not apply to our model 162. Policy insights and recommendations arisen from the modeling process were relevant for sustainable system functioning and management. 163. not aplicable (yet) 164. This item does not apply. 165. I have no experience with PIoR and cannot evaluate if a PIoR can provide innovative solution to manage the system. 166. What is the insight/recommendations? This is premature question

	15	Congruence of PIoR with culture	167. Not sure how/if this applies 168. I think this is NA because we did not test the model with 3rd parties 169. Still needs to be applied outside the group 170. It was not presented to the stakeholders. 171. Yes 172. The measures and policy insight brought up in the modeling process were agreed by those who participated, but the workshop didn't involve any stakeholders or governance representatives. 173. not aplicable (yet) 174. Yes, I totally agree with this item, valuing the creative process that took place in the workshop. 175. There is no PIoR. 176. What is the insight/recommendations? This is premature question
	16	Boundary adequacy (as policy)	177. The model only identifies policy boundaries as a potential issue in accomplish its purpose but makes no clear recommendations. 178. I think this is NA because we did not test the model with 3rd parties 179. I have doubts and lack capacity to come to a conclusion regarding this. 180. It was not tested. However, it could be done by using some weighting coefficients to the variables and prioritized them, depending on the location (depending on the strategic guidelines). 181. Yes. I believe so. A larger system would be difficult to approach because it would carry many more variables 182. A change in the scope or boundary of the model is likely to affect the policy considerations brought up in the process. If applied to a larger system, policy insights are likely to be adjusted to a more general socio-ecological scenario, as opposed to the current focus on the Macaronesian Demonstration Area. 183. not aplicable (yet) 184. I think so. The model makes an approximation by establishing relationships at the level between the different variables considered, also from a socio-cultural perspective, so such a modification would be necessary. In the case of application to a wider system, a revision would be fundamental, as it would entail a greater number of variables to consider and, therefore, a more complex process of establishing causation. 185. There is no PIoR. 186. It is already a quite large system, but is still missing out some important nations involved (Norway, EU, Russia, UK). I think it may be hard to apply to a different locations…it is quite specific 187. What is the insight/recommendations? This is premature question 188. We tested the model for the TA and so the policy recommendations are useful for this case study
	17	Learning	189. As mentioned above, the model was created by a poorly-diversified pool of stakeholders (mostly biologists) who tend to share a "common" view of the problem. 190. I think this is NA because we did not test the model with 3rd parties. But I enjoy the process.

			191. Still needs to be applied outside the group 192. It was not presented to the stakeholders. However, as participant (team member), it could be consider a 4. 193. Yes because these aspects were already within our knowledge. If we want to learn more, we know where to look. 194. Participants of the modeling process reported a satisfactory learning experience from the workshop. There was also support on ways to further learning. 195. We didn't test this model with stakeholders. 196. Yes, I think the workshop facilitated the learning, from a practical application and co-design approach. We also know where and how to expand the information, if needed. 197. I did learn new things about the socioeconomical part of the model. If I express interest to learn more, I am confident Marine Sabres project members would assist me 198. The modelling process was very informative and I was satisfied with the process. Following the modelling meeting I have since been supported in learning more about the DA
	18	Engagement	199. I think this is NA because we did not test the model with 3rd parties 200. Still needs to be applied outside the group 201. No 202. Yes, I think so in the discussions that followed the built of the model 203. Stakeholders did not participate in the modeling process, thus could not engage with the team. 204. We didn't test this model with stakeholders. 205. We have not yet had the opportunity to have a space for discussion with stakeholders following the modelling exercise. If the item is associated with the workshop participants, then the rating for this indicator would be 5, as we jointly validated the design. Likewise, after the construction of the CLD, a PESTLE analysis was carried out, identifying a variable in the diagram linked to each dimension (political, economic, social, technological, legal and environmental). 206. Stakeholders did not participate during the model exercise. 207. Hard to say, because I am not in direct contact with the stakeholders, plus we did not have the possibility to show those result to anyone yet
Administrative, review, and overview	19	Ease of Enrichment	208. The model as is and the way it was created should be easy to accommodate alterations that are deemed necessary. 209. I think this is NA because 1-we did not include a specific policy and 2-we did not use quantitative data. 210. I have doubts and lack capacity to come to a conclusion regarding this. 211. Maybe with a more specific model, for a specific question, could benefit more with an enrichment to perform new policies. 212. It depends on the new information and where it will fit, because the model was already too complex 213. New information and variables could easily be incorporated into the current model, with a relative increase in complexity. 214. I believe new information can easily be included in this model, to update or test the effects of new policies. 215. Data can be easily added. 216. It depends on the information to be incorporated and changes in the context. In this sense, the complementarity of the analysis by applying a PESTLE can be very useful. 217. I believe that can be done.

			218. It is possible, but not easily.
	20	Time & Cost of the Intervention	219. The model was created within the time expectation. 220. Yes. Yes, the biography send before the exercise was useful, as having a brief discussion within the DA team. Have at least someone from politics and economy department would have improved the exercise. 221. Yes; have people from other backgrounds. 222. Yes. Brief discussion and bibliography were essential. 223. The time to build the model was a bit too long although I understand the reason why. Actually, the moderator prevented it from being longer. I have no suggestion to improve this aspect 224. The modeling process concluded within the expected schedule, although some time was cut from the planned breaks. As a recommendation, it would be helpful to communicate the purpose of the exercise ahead of time in order to be better prepared. 225. My only recommendation would be to make the WK (if online), more inclusive, by using the tools I mentioned earlier. And I believe it would be helpful to have someone helping the session moderator to manage the speeches and the chat while the notes are being recorded. 226. It lacked the involvement of more stakeholders and more time to reach a consensus. 227. I think the time was adequate, although at the beginning I had the feeling that the definition of variables was done very quickly, which I found difficult to integrate. As for the cost, I have nothing to comment 228. We almost managed to finish the model within the allocated time but not completely. There are no recommendations to improve efficiency 229. It would have been very helpful to have had this workshop at the very beginning of the project. 230. We did all the process within the planned time 231. Yes. Th bias from mentioned before could be minimized if not only Biologists be involved in the construction of the model. Scores should be added to connections.
	21	Documentation	232. The rationale model creation is well fundamented, but the thought-process behind the choice of each parameter present in the model, although discussed, was not documented. 233. We should have record the discussion the we had while constructing the model. 234. Maybe there should have been a manual-like report. 235. A step-by step document could help to in the modeling process replicable (and/or video examples). 236. Yes 237. Unsure of whether or not the modeling process and changes along the way were documented to allow for replicability. 238. Yes, the bibliography shared previously was helpful. 239. No, I’m not aware of any document about the process of making the model. 240. Yes, I believe that the available documentation facilitated the development of the modelling exercise. 241. It is documented but I cannot say if it is satisfactory. 242. Who is writing the documentation?
	22	Replicability	243. The rationale model creation is well fundamented, but the thought-process behind the choice of each parameter present in the model, although discussed, was not documented.

			244. No, because we did not officially write a document while we were constructing the model 245. Lacks a guidebook 246. Although the written documentation is useful, for replicability will be better to have a facilitator/moderator, that would helps to understand better the construction for the model (or a short video with some examples, to allow to understand the models design). 247. I hope so 248. Unsure of whether or not the modeling process and changes along the way were documented to allow for replicability. 249. There is no written documentation. 250. Although it would be desirable, perhaps the replication process should be accompanied by training sessions to define and understand the different aspects of the model 251. I am not sure. 252. Well, there is a lot of specific expert knowledge involved (biologists/social scientists/economists/etc.)….not sure. 253. I have not seen any documentation 254. Not all the independent third parties
	23	Audit or cross-validation	255. I was involved in the model process 256. I think this is NA because we did not test the model with 3rd parties 257. Everything looks like in order 258. The model was not validated by the stakeholders 259. Model makes sense and it doesn't contradict any physical law or rigorous social norms. No, to the last question 260. The model and PIoR make sense according to the team, and no legal and social norm contradictions were detected. Likewise with previous experiences. 261. We haven't defined PIoR. 262. Unable to measure even with someone outside the modelling process. However, I respect the questions raised in this, I believe that the model makes sense and does not contradict any law or social norms in its possibility of application. 263. I was involved in the modeling process. 264. Well, I have been involved in all of this… 265. The model makes sense, has no contradictions, and is valid. However, we didn´t test it, so it is difficult to measure its adequacy.
	24	Higher-level Model review	266. I think this is NA because we did not test the model with 3rd parties 267. Still needs to be applied outside the group 268. It should be considered by present the model to decision makers. However, takin into account that those decision makers were not involved into the model process, it could be more difficult to understand the model. 269. Yes, to my opinion 270. The model built managed to fulfill the expectations and objectives of the proposed exercise. 271. I believe it will fulfill the expectations. 272. This item does not apply at the moment.

			273. I was involved in the modeling process. 274. I am still not 100% sure what the expectations are really to be honest, as we are not sure yet how exactly the results of the workshop will be used in further process 275. This has not been done
	25	Walkthroughs	276. I think this is NA because we did not test the model with 3rd parties 277. Still needs to be applied outside the group 278. The model was not presented to any other groups besides the project team members. 279. I believe this does not apply because we had no documentation 280. Despite being unsure of whether or not the modeling process and changes along the way were documented to allow for replicability, the model seemed correct according to the team. Main issues were also accounted appropriately, as well as considerations regarding policy. 281. We didn't test it and we don't have PIoR yet. 282. This item does not apply at the moment. 283. I was involved in the modeling process. 284. Hard to say, because I am not in direct contact with the stakeholders, plus we did not have the possibility to show those result to anyone yet 285. This has not been done
	26	System-improvement	286. I think this is NA because we did not test the model with 3rd parties 287. It was not performed. However, an adaptative capacity should be consider (improved by software used to build the model) to include the system changes, to updating the model. 288. We are not yet able to answer this 289. System behaviour response to policy implementation not yet tested for the model built. 290. Not yet, but we have the potential to do it. 291. This item does not apply at the moment. 292. Was not done during workshop. 293. Not sure about this 294. I agree it should be verified. It is possible to connect some changes, but this was not done yet.