MareFrame

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

Deliverable No. 7.2

Project acronym:
MareFrame

Project title:
"Co-creating Ecosystem-based Fisheries Management Solutions"

Grant agreement No: **613571**
Project co-funded by the European Commission within the Seventh Framework Programme

Start date of project: **1st January 2014**
Duration: **48 months**

Due date of deliverable:	31/12/2016
Submission date:	31/12/2016
File Name:	D7.2 MAREFRAME_Analysed case studies with respect to project progress evaluation
Revision number:	4
Document status:	Final[1]
Dissemination Level:	PU[2]

[1] Document will be a draft until it was approved by the coordinator

[2] PU: Public, PP: Restricted to other programme participants (including the Commission Services), RE: Restricted to a group specified by the consortium (including the Commission Services), CO: Confidential, only for members of the consortium (including the Commission Services)

Revision Control

Role	Name	Organisation	Date	File suffix[3]
Author/Task leader	Valerio Bartolino	SLU	16/02/2016	
Author	Alan Baudron	UNIABDN		
Author	Barbara Bauer	SU		
Author	Christopher Desjardins	UI		
Author	Bjarki Thor Elvarsson	MRI		
Author	Beth Fulton	CSIRO		
Author	Jamie Lentin	STL		
Author	Kåre Nolde Nielsen	UiT		
Author	John Pope	NRC		
Author	Margarita Rincon	CSIC		
Author	Javier Ruiz	CSIC		
Author	Gunnar Stefansson	UI		
Author	Gheorghe Sarbu	NIMRD		
Author	Mark Dickey-Collas	ICES		
Author	Jan Horbowy	NMFRI		
Author	Fallon Niall Gerard	UNIABDN		
Author	Francesco Colloca	CNR		
Author	Doug Butterworth	UCT		
WP leader	Olavur Gregersen	SYN		
Coordinator	Anna K. Daníelsdóttir	Matis	30/12/2016	AKD
Administrative Coordinator	Oddur M. Gunnarsson & Kristinn Ólafsson	Matis	29/12/2016	OMG, KÓ
Scientific Coordinator	Gunnar Stefánsson	UI	28/12/2016	GS

[3] The initials of the revising individual in capital letters

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

Deliverable D7.2

Analysed case studies with respect to project progress evaluation

28/12/2016

Executive Summary

This report is a deliverable of Work Package 7 (WP7 – Synthesis & training development) of the FP7 MareFrame research project. The report comprises two main parts. The *first* part presents a protocol for the comparison and evaluation of the ecosystem models that have been developed and applied within the MareFrame project. This is to inform on their general suitability to provide support to an Ecosystem Approach to Fisheries Management (EAFM) (task 7.2.1). The *second* part of the report reviews and evaluates the progress on the ecosystem models developed for the various case studies (task 7.2.2). This work is based on application of the protocol developed in the first part, with the addition of a procedure to assess progress in the application of the models. The timing of this deliverable (M36) was specified to provide a suitable checkpoint to assess status, as well as to provide specific recommendations on corrective measures where needed, so as to ensure that the ecosystem modelling- related objectives of the MareFrame project will be achieved.

This report should be viewed in the context of the subsequent report D7.5, which will evaluate the Decision Support Framework (DSF) developed within the project and applied in the various case studies. Within the MareFrame project, use of the term DSF encompasses the ecosystem models, decision tools and co-creation process developed to facilitate progress towards an EAFM. The scope of this report is limited to address the comparison and use of the ecosystem models that comprise the empirical basis for the DSF, and hence considers only scenarios relevant for the evaluation of management alternatives. Furthermore, this report is tightly linked to activities in other parts of the project. In particular, procedures for model comparison will be developed in concert with activities in WP4 (especially with deliverable D4.2), which has established common reporting procedures for model outputs. The outcomes of the comparative analysis will inform WP5 and WP6 in particular.

Ecosystem models show considerable variability in their output for all the case studies, which is a consequence of the high structural uncertainty inherent in ecosystem models. Some of these differences arise from the range in scope (from tactical to strategic) covered by the different models examined, as well as from the different extents to which they focused on securing good fits to the data available. In general, comparative approaches are recommended as the way forward, both to quantify structural uncertainty and to find results which are robust to model formulation.

All of the MareFrame case studies adopted the co-creation approach. This led to confirmation of the high potential which ecosystem models possess to highlight the trade-offs to which fisheries management needs to give attention. However, in several cases the approach also served to emphasise the limits of the current models and the difficulties in implementing them to address some of the specific '*co-created management objectives*'.

A general feature of the ecosystem models considered is that increased model complexity comes at the expense of precision and ability to fit available data. The inclusion of more species, trophic layers and processes often requires more assumptions, readily finds itself compromised by paucity of data, and can lead to difficulties in achieving statistically appropriate fits to data. Nevertheless, the management questions posed, as well as the development of the associated decision support tools, for the different case studies were found to require models which addressed certain aspects of this increased ecosystem complexity. These aspects go far beyond what traditionally needs to be considered for single species stock assessments.

Regardless of the fact that a management question may require the provision of analyses to inform a short term tactical decision, it is the long-term implications of that decision for the ecosystem are of most interest in an EAFM context. In several of the case studies, rather than indications of which model outperformed the others, what emerged was the need for complementarity. Management of fisheries requires explicit recognition of the complexity of individual fish populations in terms of their abundance and demographic structure, but this does impose strong limitations in the context of an EAFM unless this is limited to a handful of the most important targeted species.

There is a particularly strong need for methodologies which synthesise the considerable quantity of outputs generated from multiple ecosystem models. This is especially the case in a framework which involves close interactions with stakeholders, as applies in all the MareFrame case studies. The DST has an important role to play in this. In the spirit underlying MareFrame, the DST should be able to incorporate output from multiple models as well as take account of some other sources of uncertainty, such as arise from stochastic aspects and from environmental variability. Caution is recommended in the combination of multi-model outputs into integrated statistics, and simple model averaging should not replace an in-depth understanding of these uncertainties so that they can be accorded appropriate relative weightings in advising decisions.

Contents

Introduction 7

Comparison and evaluation of ecosystem models with regard to their relevance for an EAFM 8

- Data consistency for consistent comparison 15
- Model parametrization and forecasting 15
- Metrics and GES indicators for comparison 16
- Uncertainty 19
- MareFrame models for comparison 20
  - Gadget 20
  - EwE 21
  - Other models 22
  - Atlantis 23
- Comparing models in practice 24

Evaluation and feedback regarding the application of models in case studies 25

- Baltic Sea case study 26
- North Sea case study 31
- Icelandic case study 39
- West of Scotland case study 41
- Iberian case study 44
- Mediterranean case study 50
- Black Sea case study 53

Conclusions 61

References 63

Introduction

This document reports on the development of a comparison and evaluation protocol, and then its application to a number of case studies. The aim is to contribute to the overarching objective of the MareFrame project, which is to remove barriers that prevent a more widespread use of an Ecosystem Approach to Fisheries Management (EAFM). Ecosystem models represent key instruments to achieve EAFM. The MareFrame project proposal identified several such barriers which were directly related to the development and application of ecosystem models; it also indicated ways that the project would address overcoming or alleviating these barriers (see Table 1).

Table 1. The most relevant modelling-related barriers for adopting an EAFM, with summaries of how MareFrame intends to overcome or alleviate them.

Barrier to acceptance and use of EAFM	How MareFrame deals with this barrier
Lack of consensus on what models to use and what they should contain	Include different models and different schools of modelling in the project, evaluate respective applicability, strengths and weaknesses
Lack of guidelines with respect to what models and tools are suitable in what circumstances	Use and further develop "best practice" guidelines based on experience from running different models in different cases (FAO, 2008)
Lack of detailed knowledge relating to population structure, spawning components and trophic levels	Include novel analytical data to develop new ecological knowledge
Existing data are fragmented and in various formats, data still missing, particularly environmental and socio-economic	Develop a uniform and harmonised database designed to support ecosystem based modelling , integrate environmental and socio-economic data into the database
Existing ecosystem models do not not explicitly provide stock assessment and management advice	Extend existing ecosystem models and assessment methods to meet the needs of managers and operators

The primary purpose of this report is to contribute to the goal of overcoming ecosystem modelling-related barriers in the ways suggested in this Table.

In response to the first two barriers mentioned, the *first* part of this report presents a comparison of the ecosystem models that have been developed and applied within the MareFrame project, and evaluates their general suitability to support an EAFM. As MareFrame is a European research project, funded by the 7th Framework programme, the term EAFM should be understood mainly in relation to current European policy frameworks within the context of the management of marine fisheries environmental management. The main policy frameworks of relevance in this context are the reformed Common Fisheries Policy (CFP - EU 2013) and the Marine Strategy Framework Directive (MSFD - CEC 2008). MareFrame has developed a protocol (task 7.2.1) specifically for this purpose. It should be noted, however, that two of the case studies (the Iberian and the Icelandic studies) relate to fisheries which do not fall under the CFP; moreover, the latter is not subject to the MSFD. For these case studies, an EAFM is informed by national policies. The protocol focuses initial attention on the problem of achieving consistent treatment of model input data across different ecosystem models to the extent possible. The MareFrame approach to address this is presented briefly. It relates primarily to constructing a database to facilitate uniform data treatment for different model runs and across different ecosystem models. This approach addresses the fourth barrier listed above.

It is of particular relevance to the second and to the last barrier listed that the MareFrame project is mainly an applied project, which seeks to contribute to the development of tools and processes which support an EAFM through case studies. Accordingly, the *second* part of this report is to evaluate progress of the application of ecosystem models for the case studies (task 7.2.2). It does this based on a procedure developed for this purpose. The more specific objectives that inform this evaluation are those of WP5:

a) to implement the ecosystem-based modelling approaches developed in WP4 in seven different marine ecosystems within the European Union; and
b) to investigate the effect of fishing and different climate scenarios on key ecosystem processes, and to provide a basis for the development of decision support tools in WP6.

This evaluation, which is due one year before the project terminates, has the intention to offer a basis to provide feedback and advice at a case study level; this is to ensure that the objectives related to model development in each case study are achieved.

The report is divided into two main parts.

- A protocol for the comparison and evaluation of ecosystem models with respect to their suitability for fisheries and environmental management purposes, and to their ability to predict the responses of a multispecies community of fish to changes in fishing mortality (task 7.2.1). After an introduction to the issue of such a comparison and evaluation for the purposes of an EAFM, the report details key aspects of the comparison process including issues related to data consistency, model parametrisation and uncertainty. Wherever possible it also suggests a MareFrame-related approach to these issues given the array of ecosystem models considered in the project. The protocol includes a list of indicators covering fisheries, ecosystem, economic and social aspects, and provides practical advice on how to perform model comparisons.
- An analysis of progress on the ecosystem models under consideration, including feedback on model development, at a case study level (task 7.2.2). Model comparison is based on the protocol developed in the first part of the report, and is taken forward by assessing the models within each case study against a common checklist.

Comparison and evaluation of ecosystem models with regard to their relevance for an EAFM

The ecosystem models applied within MareFrame case studies will be compared and evaluated with respect to their capacity to support an ecosystem approach to fisheries management (EAFM). This capacity not only relates to the ability of models to predict responses of a multispecies community of fish to changes in fishing mortality but also to predict impacts of fishing on other environmental indicators. In addition, the relevance of the models may depend on their ability to estimate changes in indicators relevant for assessing economic and social impacts of changes in management approaches or environmental change.

Most of the MareFrame case studies are subjected to EU policy frameworks, which comprise a specific commitment to pursue an ecosystem approach. The most important polices in this context are the CFP and the MSFD, which both comprise a commitment to an ecosystem approach. In order to ensure its relevance, it makes sense to consider the comparison of the ecosystem models in relation to the way EAF is defined in these polices.

"The CFP shall implement the ecosystem-based approach to fisheries management so as to ensure that negative impacts of fishing activities on the marine ecosystem are minimised, and shall endeavour to ensure that aquaculture and fisheries activities avoid the degradation of the marine environment" (EU 2013).

"Marine strategies shall apply an ecosystem-based approach to the management of human activities, ensuring that the collective pressure of such activities is kept within levels compatible with the achievement of good environmental status and that the capacity of marine ecosystems to respond to human-induced changes is not compromised, while enabling the sustainable use of marine goods and services by present and future generations" (CEC 2008).

Motivation

The main motivation for using ecosystem models in management arises from the existence of trade-offs among different management objectives, which is a central aspect of the ecosystem approach to fisheries management (EAFM). One of the numerous benefits of an EAFM is that it explicitly considers trade-offs between different stakeholder priorities, balancing human (i.e., socio-economic) and ecological needs (Garcia et al. 2003). Ecosystem models provide qualitative and quantitative estimates of the expected benefits, costs, and risks associated with alternative management actions. Thus, they aid the understanding and quantification of that difficult trade-offs, even in complex contexts, such as in multispecies fisheries (Hilborn et al., 2004; Link, 2010; Forrest et al., 2015). Therefore, the development of appropriate metrics which aid the interpretation of model outputs in terms of trade-offs is a crucial element of a protocol for ecosystem models comparison.

The quantification of trade-offs serves as a useful common denominator for analysing results across the MareFrame case studies, considering the diversity of modelling approaches and management issues within this project. In the Common Fishery Policy (CFP) area the implementation of multiannual management plans (EU 2013) based on MSY targets is agreed as a way forward for the management of the European fisheries. However, it is recognised that in a multispecies context identification of MSY targets may not always be possible. This implies that a qualification and quantification of the trade-offs among fisheries targeting different species is necessary to address the issue. For instance, in the Baltic Sea case study a trade-off is represented by how different fishing levels on the cod, sprat and herring stocks will impact the yields of the bottom trawl and the small pelagic fisheries. In the Mediterranean case study rebuilding of the hake stock based on single species MSY targets is like to generate trade-offs with the management of the sardine and anchovy stocks. In the Chatham Rise case study, the potential development of seabed mining of phosphorite is likely to create trade-offs with fisheries and conservation interests in the area. In the West coast of Scotland case study, the objective to seek recovery of cod and whiting stocks is likely to invite trade-offs with regard to the economic performance of demersal fisheries.

For this reason, we propose a number of metrics and visualisation procedures for comparative analysis of the ecosystem models implemented in MareFrame, which could help to explicitly characterise the trade-offs between different ecosystem components, or attributes, specific to the different case studies.

Before proposing a protocol for model comparison, however, it is important to briefly review the role of ecosystem models in supporting the emerging needs of ecosystem-scale assessment and to identify some of the boundaries of the models and questions of interest for MareFrame for which model comparison is needed. Real applications of multispecies models for the assessment of marine resources started during the 1980s and 1990s (Andersen and Ursin 1977; Laevastu and Livingston 1980), but for almost two decades they have rarely influenced the management of marine resources, and typically have been, at best, used to inform on a qualitative base ecosystem considerations for single species assessment. Although there are exceptions – such as the use of an ecosystem model to help shape broad scale management practices in the Australian Southern and Eastern Scalefish and Shark Fishery (Fulton et al. 2014).

Starting with the 1995 FAO Code of Conduct for Responsible Fisheries and following with the 2001 Reykjavik Declaration on Responsible Fisheries, there has been an incremental increase in interest around the development of operational approaches for the implementation of an EAFM. The development of complex multispecies (and full ecosystem) models has served such demand by enabling scientists to quantify the impacts of fishing and other human activities on the status and function of ecosystems (Francis et al. 2007; Levin et al. 2009). Considerable work has been done to classify the current range of multispecies and ecosystem models (e.g. Hollowed et al. 2011; Fulton et al. 2003; FAO 2008; Plaganyi et al., 2014). Ecosystem models and their applications have been categorized as either (i) conceptual (aimed at developing a qualitative understanding of ecosystem processes); (ii) strategic (focused on broad scale assessment of potential directions and patterns of change); or (iii) tactical (directed at supporting specific management decisions) (FAO 2008). Some model types can be sensibly used across all 3 use types, but other models are best suited to one role or another. For example full system models such as Atlantis are best used for startegic purposes or, in some cases, conceptual system understanding, but are inappropriate for use as tacitical quota setting tools.

The modelling capacity of the array of models available to MareFrame spans all these three model use-type categories. However, the primary objective of using ecosystem models in the project is *'predicting change of multi-species fish communities to fishing pressure and environmental/climate effects'* and *'supporting fisheries management while considering broader environmental and ecological influences'*, which points to a higher importance of strategic use of models over the other types of uses. While a large part of the fisheries management still involves tactical decisions, arguably it is the medium- and long-term implications of such decisions that matters in terms of human impacts on the structure, function and resilience of ecosystems (Francis et al., 2007) and the sustainability of fisheries.

Increasing demand for the development and application of multispecies (and ecosystem) models arises from the realization of the potential importance of taking species interactions into account when managing fisheries since ignoring these, as is done by single species models, may lead to unrealistic predictions. Given the interconnected nature of ecosystems, consideration of biological interactions, such as predator-prey interactions, in long-term management decisions may substantially change the perspective and the way we use and manage marine resources, both at the level of individual fish populations and entire marine ecosystems (Stefansson, 2001).

Increasing model realism by modeling biological interactions generally comes at the price of increased model complexity. Ecosystem models typically require input information on a large number of ecosystem components and processes, and also a broader spectrum of expertise including fishery science, fish biology, oceanography, mathematics, statistics, economy and computer science. This has stimulated the need for multidisciplinary expertise and team work when operating complex multispecies models. However, the details of complex models are (typically) only understood by experts, therefore, it is their task to rigorously compare model results to estimate the uncertainties of their conclusions. Experts and decision-makers then are able to work together to distill model results for their inclusion in decision support tools based on the specific needs of stakeholders.

Model inter-comparison

Determination of appropriate methods for comparing different ecosytem modelling approaches is a highly active area of current research (e.g. Plaganyi 2007). The number of ecosystem models available to scientists has increased exponentially during the last two decades (especially as easy access to computing resources has expanded). Their increased usability and the availability of datasets for different ecosystem components has made complex ecosystem models significantly more accessible, as demonstrated by the number of trophic and system-wide ecosystem models implemented around the world (Fulton 2010). The interest and uptake of ecosystem modelling applications has been stimulated by a growing demand for more ecosystem-integrated fisheries management advice. Model comparison is necessary in any attempt to maximise the complementary strengths of different approaches, to answer specific management questions, including their potential limitations and un-intended consequences. For instance, ICES adopted the concept of using a model 'key-run' for reporting and comparing results from multispecies models in a transparent way (WGSAM 2009, 2010). However, the procedures for implementing and comparing key-runs are still only partly integrated among different modeling frameworks, and ICES has mostly model specific key-run procedures (i.e., EwE key-run). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) has developed a simulation protocol to bring together and compare a large number of impact models across sectors and scales to forecast the impacts of different levels of global warming (Tittensor et al in prep). While this has dealt with (and attempted to formalise) many model intercomparison issues there is still scope for refinement – especially in the form of extension when dealing with specific questions or contexts. Moreover, while it has perhaps been the most formal attempt to date to define model inter-comparison protocols in a fisheries related context, it is not the first inter-comparison effort, with the approach already used to explore questions around EAFM and related fisheries management rules – for example the sustainable exploitation of forage fish (Smith et al 2011) and ecosystem indicator performance (Fulton *et al.* 2005).

The Marine Strategy Framework Directive (MSFD) and its associated decision (EU, 2008, 2010) lists a collection of pressures and states (descriptors) that need to be assessed in terms of Good Environmental Status (GES). Each member state of the EU is expected to provide its own definition of GES, although the member states are encouraged to determine these regionally. Prescribed guidelines for their definition and assessment are provided for some, and not for others (e.g Descriptor 3 commercial fisheries provides guidance that GES is when the stocks are fished below Fmsy, whereas Descriptor 4 foodwebs provides the generic phase "occur at normal abundance and diversity and levels capable of ensuring the long-term abundance of the species"). There are described approaches

for determining reference levels for time series when a clear pressure state relationship is evident (Samhouri et al 2010, Figure 1). However, the world is not that simple.
There is now a prosed revision of the MSFD, and it shifts to include a threshold approach after the realisation that the initial MSFD assumed that clear pressure state relationships could be determined for all descriptors. However, the wants of society (i.e. policy development) was further advanced than science could offer. Often there is insufficient evidence to define targets and provide a formal state assessment and/or the links from the state to anthropogenic pressures are either weak or not sufficiently understood to underpin specific management advice (Shephard et al 2016).

Thus there is a proposed revision of the decision (EU 2010), which is currently undergoing public consultation and may be agreed in 2017. In this revision the mechanisms to estimate GES for each descriptor are more prescribed, with many now requiring specific metrics (e.g. for foodwebs the abundance and size structure of a minimum of three trophic guilds needs to be monitored). When it is difficult to determine the reference point of a stressed system (Figure 1A) or if no relationship is known to exists relating to a pressure, or if the prevailing natural conditions have a bigger impact on the ecosystem state than anthropogenic pressures, then the concept of thresholds can be used Thresholds are the bounds or limits that are placed on each metric used when assessing GES. The thresholds can be a minimum value, maximum value, or set of bounds, for the metrics, beyond which the monitored metrics should not pass. If the ecosystem state passes those thresholds then action should be triggered; an evaluation of the pressures that impact that monitoring metric, further research or direct action to reduce pressures. This concept is very similar to that described by Shephard et al (2016, Figures 2 and Figure 3) where surveillance indicators and bounds are described, similar to metrics for criteria and thresholds. The revision of the decision will ask for member states to set these thresholds by region or subregion (such as for biodiversity), or EU wide (such as for litter and noise).

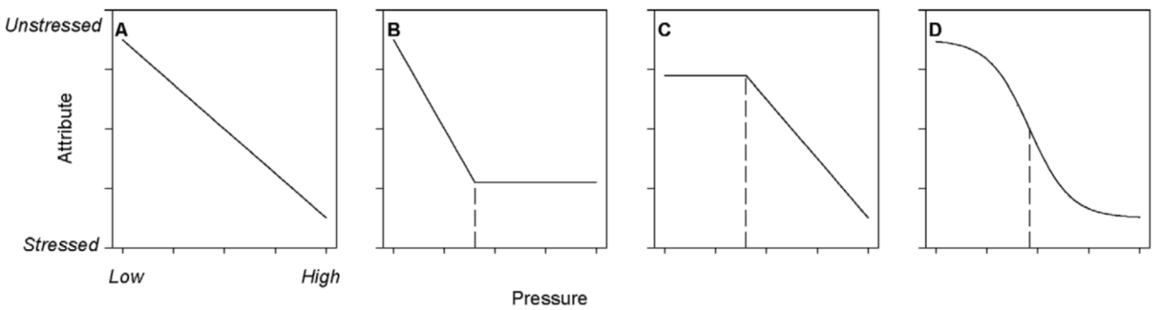

Figure 1: Relationships between hypothetical ecosystem attributes and anthropogenic pressures. Attribute values range from unstressed to stressed (and the levels of the pressures applied have been scaled relative to a theoretical maximum. A utility threshold cannot be defined objectively for the linear model (a), but can be defined objectively for the two piecewise models (b and c) and the sigmoidal model (d). Equations for the models and the location of the utility thresholds. In (b-d), the threshold pressure is indicated by the dashed lines. (Samhouri et al 2010).

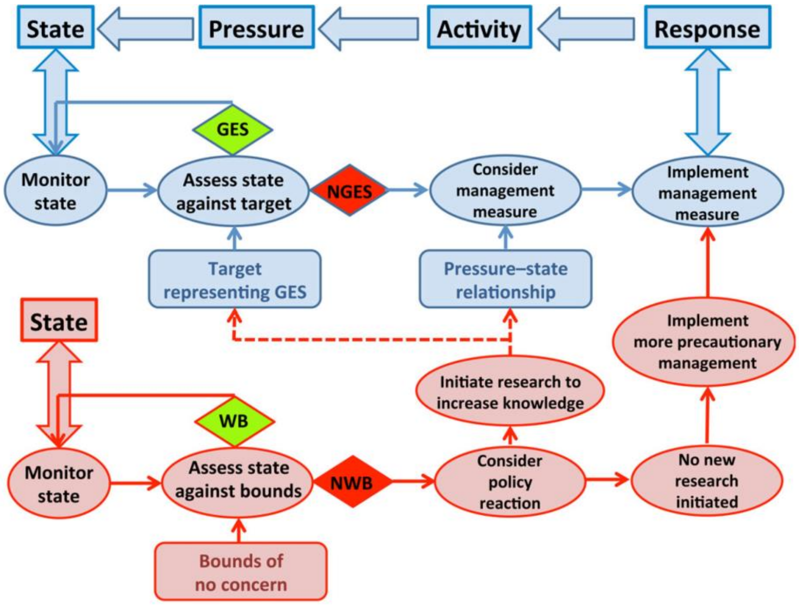

Figure 2: Diagram illustrating how surveillance indicators (red process) can complement operational indicators (blue process) in an Activity Pressure State Response (APSR) approach to the MSFD. Operational indicators evaluate whether state is meeting (GES) or failing (NGES) "GES" targets. Surveillance indicators evaluate whether state is within bounds (WB) or not within bounds (NWB), where these bounds represent the upper and lower limits of a range in state for which there is no "specific cause for concern" (Shephard et al., 2016).

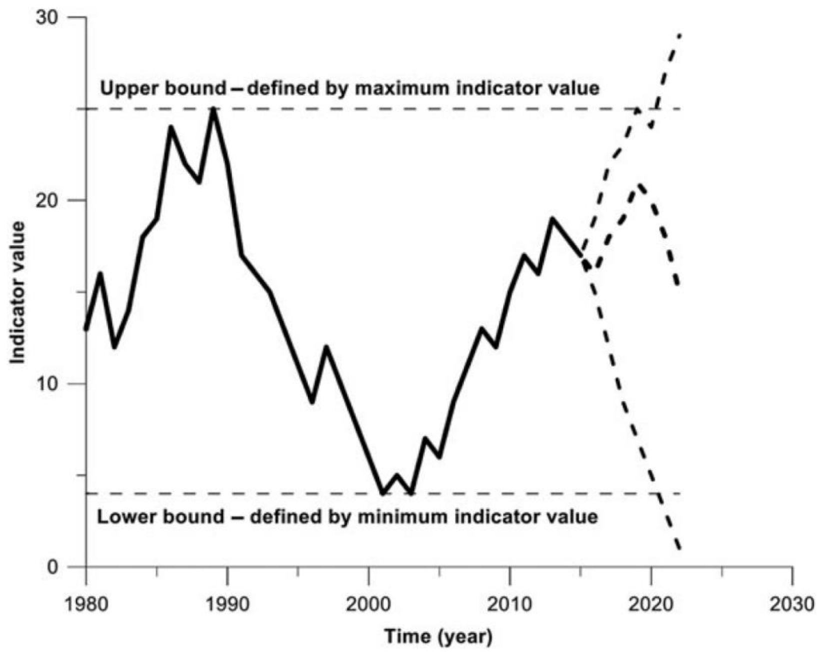

Figure 3: Schematic of a generic surveillance indicator time-series (solid black line) showing historical upper and lower bounds within which the indicator has varied over the time-series duration (fine dashed black lines). Future variation in indicator values beyond these bounds (e.g. increasing and decreasing dashed black trajectories) (or thresholds) implies that the ecosystem component in question is changing towards a state not previously experienced. Such a situation would represent a "specific cause for concern" and should trigger a policy reaction since, under these circumstances, the knowledge necessary to underpin reliable scientific advice would increasingly be in short supply (Shephard et al., 2016).

Currently the thresholds are expected to be set against empirical monitoring and observations, however nothing in the decision states that synthesised and modelled time series cannot be used. Also ecosystem models can be used to explore the acceptable levels of the thresholds. In addition, modelling can explore the likely consequences of actions and the need to consider any trade-offs when proposals are made to change management of human activities. These are called management strategy evaluations, similar to that described in Rijnsdorp et al. (2012) where the consequences of closing flatfish spawning areas are examined in terms of impact on fishery profits, biodiversity and sea floor integrity. The MSFD descriptors are a mixture of pressure and state and when using models to explore the relationship and the setting of thresholds this mix of pressures and state can be better accounted for than using empirical information alone.

The challenge of developing an effective protocol for model comparison is that it should be sufficiently general to be adopted by all the case studies. Nevertheless, at the same time it should represent an operative procedure able to be adapted to the models' primary objectives within each case study, especially considering that ecosystem models in MareFrame are context- and question-driven.

Within all the case studies we will have the two-fold task of comparing results from both competing models implemented in the same modelling framework (i.e., structurally-similar models such as alternative Gadget formulations) and models implemented in different frameworks (i.e., two structurally-distinct ecosystem models such as Gadget and EwE, Atlantis and EwE or others). In the first case, standard procedures exist to evaluate model performance which could form the basis of comparisons (see the section 'MareFrame models at comparison'). In contrast to the familiarity of comparing models within the same framework (which is the foundation of standard sensitivity analyses), there are far fewer published examples of comparing structurally-distinct ecosystem models (e.g., Fulton and Smith, 2004; Kaplan et al., 2013; Forrest et al., 2015). Such comparisons are based around considering the dynamics of specific state variables either in comparison to a common observation time stream or in response to the same drivers. It is this later approach that will be used in this case. That is the comparison of results will be done in terms of response to specific management options and the results of the comparison used to test whether robust advice can be developed and to what degree they may be trusted if quantitative results differ among ecosystem models (Collie et al., 2014).

The first element of our model comparison framework is the use of a common database to ensure consistency and transparency in terms of the resolution and sources of input data and validation data. Second, compared models are parameterized independently of each other. Finally, we established a set of common metrics that are related to trade-offs involved in management and can be computed by all models

In addition, the MareFrame model comparison differs from previous studies also in the use of operating models in a subset of the work. In the specific, Atlantis or minimum realistic models (MRM) will be used to generate input data for other models so that the ability of the other models to cope with specific data issues and ecosystem responses can be assessed across a spectrum of data availability (i.e. data rich and data poor) situations. This kind of simulation testing of assessment methods (using a management strategy evaluation approach) has been performed for single species assessments previously (Rademeyer et al. 2008), but is rare at multispecies and ecosystem scales (see Howell and Bogstad 2010 for an exception) and sets MareFrame apart.

Data consistency for consistent comparison

An important area of improvement for comparison and validation of ecosystem models is the application of procedures and tools that guarantee consistency in the use of input data. Formulating model input files by hand involves risk, i.e. procedures to transform data into a form suitable for each model may not use data in an identical fashion. Worse still, as new data becomes available, it may be added to one model but not necessarily to another.

Instead, we have automated as much of this work as possible, using a database to store, collate and transform data into model input files. Functions to handle this database are available as an R package (mfdb), which can be installed following the instructions at https://github.com/mareframe/mfdb. This database allows for the automation of the compilation of model files and the aggregation of data at the appropriate level for the models. So, instead of models being produced through a series of manual steps, an R script can be used that is able to generate input for more than one model type. This makes the work of a modeller much faster, as tweaking the way data is collated is no longer a time-consuming manual process. In addition, it means we can update multiple model input files in lockstep, making it much harder to, for example, update sample aggregations over area without updating the area definitions themselves. Finally, multiple models can be updated at the same time, meaning it's much less likely that one model accounts for new data whereas the other does not.

In addition to streamlining the model development process, the database system allows for direct comparison between fundamentally different modelling approaches. This can be implemented through a special bootstrap procedure, described in Elvarsson et. al (2014), specifically designed to handle disparate datasets used in multi-species models. The procedure is based on spatial units that are resampled to produce replicate datasets used in the model fitting phase. Models can then be compared based on various output statistics with uncertainties.

An additional area of model comparison which will be addressed in MareFrame consists of the use of Atlantis or MRM operating models – those are models representing "truth" from which data is generated and other models are fit with uncertainty generated to mimick reality. The benefit of this approach is that because "truth" is known the capacity of the other models to reflect that state and how that capacity differs with data availability or system state can be directly assessed. In this context, in MareFrame, a MRM will be used to compare the performance of Gadget and other models in South Western Waters, Gulf of Cadiz case study while Atlantis will be used to compare the performance of Gadget and EwE in Iceland and the Strait of Siciliy and it will be used to generate data for a data poor ecosystem. Combinations of different modelling scenarios will result in selected outputs (e.g. biomass and landings data) that will be used to create an input dataset that will span a range of 'challenges' for some of the other ecosystem models used in MareFrame. The intention behind this range of 'challenges' is not to trick the ecosystem models through using unrealistic scenarios that cause one model or both to fail, but instead to understand and to quantify the extent to which decisions converge and diverge between the different models using realistic ecosystems. The same database will be used to import selected models-generated datasets and extract them for the other ecosystem models.

Model parametrization and forecasting

Given their different structure, the models included in the comparison have to be independently parametrized with each specific calibration routine, but with consistent input data. In contrast with the hypothetical 'retrospective forecasting' approach proposed for illustrative purposes by Forrest et al. (2015), we aim for an operative evaluation of management scenarios by comparing models on actual medium- and long-term forecasts.

Forecasting should allow evaluation of trade-offs in concert with associated uncertainties. For each model in the comparison different sources of uncertainty should be taken into account to provide probability envelopes for future trajectories, rather than simply presenting single best estimate trajectories. One of the main, unique values of the MareFrame project is that it enables us to disentangle different sources of uncertainty by facilitating the comparison of both the performance and forecasts from different models within and across various ecosystems.

The actual comparisons must be based on values which can reasonably be compared. For example, most reference points cannot be compared across models (cf Fmsy, Fmax etc) but the relative status of current fishing mortality compared to such reference points can be compared. Similarly, MSY can be compared across models.

Metrics and GES indicators for comparison

Given the multitude and complexity of processes regulating the dynamics of marine ecosystems, it is unlikely that individual metrics could adequately disentangle and represent both short and long-term impacts of fishing pressure, or changes in productivity or climate (Longo et al. 2015). For this reason proposed predictive frameworks generally call for suites of complementary indicators (Rochet et al. 2005, Niemeijer and de Groot 2008). Results from simulation studies in southeastern Australia suggest that it is necessary to simultaneously consider a suite of indicators spanning a wide range of processes and biological groups if there is to be robust detection of the full range of effects of fishing (Fulton et al. 2005). As a general principle, Fulton et al. (2005) suggested that detecting signals and characterising an ecosystem would be best achieved by monitoring species groups from the following three categories: groups with fast turnover (i.e., "early warning" indicators), groups targeted by fisheries (i.e., groups of significant human interest), and charismatic or vulnerable groups (often dominated by slow population dynamics). Simple community-level indicators based on the ratio of large groups (i.e., forage species:top predators, piscivores:planktivores, pelagic:demersal) have also been shown to be effective in describing major shifts in the trophic structure of the system. However, it is worth noting that indicators reflecting broader system structure and function remain of great conceptual interest, with no simple candiates for such indicators currently available. There has been some attention put into such indicators in other monitoring and model comparison exercises, such as INDISEAS (who considered the Large Fish Index) and the LENFEST consideration of the effects of forage fish exploitation (Essington and Plaganyi 2014).

The ecosystem models used in MareFrame vary greatly in their structures, resolutions, and outputs. A review of their similarities and differences together with a list of their specific outputs is found in D4.2. Population biomass, abundance and fisheries yields have been identified among the common outputs produced by all the models and will be used to calculate a number of indicators over selected species and biological groups to link model outputs to what is defined in MSFD as 'Good Ecological Status'. To embrace the complexity of pressures and states characterizing ecosystems at a population and a community level, a wide suite of indicators has been selected as performance metrics for model comparison in MareFrame. Selected metrics span from indicators based on biological reference points, biomass based indicators, ratio indicators, biodiversity indicators, trophodynamic indicators and size-based indicators.

Table 2: List of potential indicators for ecosystem model evaluation

Biological Reference Point based indicators			
Number overfished stocks	number of stocks with SSB < MSY Btrigger, where MSY Btrigger marks the lowest boundary associated with SSBMSY	D3.2	ICES, 2015
Biomass based indicators			
Biomass			
Abundance			
Catch			
Fishing revenue			
Fishing mortality			
Fmsy			
TAC	TAC=Mean Biomass of the last n years*Fmsy		
Ratios indicators			
Pelagic:Demersal			de Leiva Moreno et al., 2000
TopPredators:ForageFish			
Piscivores:Planktivores			
Catch:Biomass			
Biodiversity indicators			
Shannon's diversity index based on landings	biodiversity index based on the proportion of species in the landings		Shannon, 1948
Number of species with landings > minimum level	landings higher than a minimum level (to be set for all models/ecosystem to be compared)		Gascuel et al., 2014

Trophic indicators			
Mean Trophic Level (MTL)	based on the mean trophic level from Fishbase (www.fishbase.org) and the weight (biomass) of species		Pauly et al., 1998
Marine Trophic Index (MTI)	MTL of predatory fish i.e. species with a trophic level of 3.25 or higher		Pauly and Watson, 2005
Size-based indicators			
Large Fish Indicator (LFI)		D3.3	
Mean Maximum Length (MML)	based on maximum asymptotic length L_∞ from Fishbase (www.fishbase.org) and the weight (biomass) of species	D3.3	ICES, 2009
Biomass size spectra		D3.3	
Socio-Economic indicators			
Revenue			
Gross value added (GVA)			
Labour cost			
Full time employment			

Many of the metrics suggested replicate the indicators currently being investigated by OSPAR and HELCOM for their upcoming assessments of GES (the Intermediate Assessment and HOLAS II respectively). Whilst not directly linking to the MSFD descriptors, the output of the size of individual in a population, the estimation of LFI and MML are useful in terms of ecosystem wide assessments. The MSFD is a mechanism for making the ecosystem approach operational, and currently the silo approach of descriptors and criteria does not easily allow consideration of ecosystem assessments. However the spirit of D1.7 (soon to be revised) and D4 is to consider ecosystem structure and function. This can really only be done by synthesising information through ecosystem models.

The ecosystem models adopted in MareFrame are mostly biophysical models lacking explicit integration of socio-economic aspects. However, for comparison of different fishing strategies economic considerations cannot be disregarded, since they are often of key policy interest and because they may lead to a shift in the composition of the catch and potentially its economic value. In this case comparison of the regimes cannot be simply based on biological yields, but it has to include also costs and incomes. Therefore, revenues and costs of management strategies will be used as metrics to compare forecasts from the socio-economic point of view. As all models will generate landings time series it is a relatively straightforward exercise to generate associated economic metrics, even if this does not then dynamically influence effort decisions made within the model (as it would if the human dimensions of the models were fully dynamic).
Considering the diversity of management issues and strategies under evaluation in the different case studies, it is likely and recommended that higher relevance will be given to those indicators more directly connected to the management issue addressed in each case study.

Uncertainty

Implementation of a precautionary approach in fisheries management requires quantification of uncertainty and risk associated with alternative management options (Harwood and Stokes 2003). Nowadays, most modern single species stock assessments routinely provide estimations of uncertainty. However, dealing with uncertainties becomes increasingly problematic as model complexity increases (McElhany et al. 2010). The issue of uncertainty affects models comparisons in a range of different ways.

First, earlier estimates of uncertainty have commonly been underestimates (Elvarsson, 2014), and an important purpose with using an operating model is to see whether a model being tested gives uncertainty estimates which reflect the actual differences from the “true” model.

Second, different models may well be affected in totally different ways by uncertainty in data (noise). This is known from single species assessments: One model may be very robust to highly variable age determinations whereas another may completely fail if age sampling is poor. The issue becomes much more difficult in the ecosystem models considered in MareFrame.

Third, different models have a different capacity to reflect natural variation in the environment (signal). Thus only a multispecies model will reflect the increase in the mortality of a prey due to an increase in the abundance of the predator. Similarly, only a handful of models can reflect environmental changes, e.g. in recruitment or growth. In the case studies where specific issues are of importance, tests need to be undertaken to evaluate whether important environmental changes are reflected in model output. The environmental variation must then also be projected into the future when comparing predictions across models.

Finally, the different models may also predict different responses to human activities, including management strategies or management plans. If several ecosystem models are available for a case study where a change in management strategy is being considered, then these need to be compared using a traditional management strategy evaluation, where one part of the MSE is the model (which might even be a single species assessment model). Naturally, such an evaluation must include plausible future environmental changes.

The resulting estimates of uncertainty can be considered when estimating costs and benefits of management strategies in the decision support tools either using the Multi-Criteria Analysis or Bayesian Belief Networks.

MareFrame models for comparison

Gadget

Gadget is a shorthand for the "Globally applicable Area Disaggregated General Ecosystem Toolbox", which is a statistical model for marine ecosystems (an earlier version was known as BORMICON - Stefansson and Palsson 1997). Gadget is an age-length structured forward-simulation modelling framework, which allows the coupling of a model with an extensive set of data comparison and optimisation routines. Processes are generally modelled as dependent on length, but age is also tracked in the models, and data can be compared on either a length and/or age scale. Models can designed as a multi-area, multi-fleet model, capable of including predation and mixed fisheries issues; however, it has also been used on a single species basis to provide tactical advice. Gadget models can be both very data- and computationally- intensive, with optimisation in particular taking a large amount of time.

There is a clear division between the modelled population and the observational data within Gadget models. Typically, once the different processes determining the dynamics of one or more populations are specified and initial parameter values, a simulation is run which results in a unique realization of the ecosystem. Only then does Gadget compare various aspects of the modelled populations with observational data, ultimately producing numeric likelihood scores that measure how well the model matched each data set. The program also computes a single overall likelihood score representing the 'goodness of fit' between the simulation and the data. Various optimization algorithms are used to estimate the combination of parameters with the associated lowest overall likelihood score.

The Gadget framework has proved to be particularly informative for the evaluation of alternative model structures (i.e., sub-stock components, spatial structures, mixed fisheries), and within a model structure, for the evaluation of how the data affect the optimization. When the model structure is a better representation of the population structure, the spatial aggregation of the likelihood component data is of less importance. Where the model structure does not adequately describe the population structure the format of the likelihood data affects the ability of the model to optimize. Consideration of these factors is part of the model evaluation process. For example, a model might be able to fit to length distributions aggregated over the entire space but not to the spatially disaggregated length distributions. The distribution of abundance could be correct, but not the size structure (Taylor 2011).

Once a biologically plausible and internally consistent model is identified, a range of alternative structurally different models is usually compared and tested on the same data. Comparison is evaluated on the ability of the models to fit to the data by using the summary score and the likelihood score of individual components (Taylor et al. 2007). Even likelihood datasets not used in the optimization can be used to compare models (Taylor 2011).

EwE

Ecopath with Ecosim (EwE) models are trophic interaction focused ecosystem models representing species or functional groups (as biomass pools) and fisheries. They are implemented in a user-friendly software that contains a set of tools to aid model construction and analysis. An Ecopath model is a mass-balanced model of an ecosystem where biomass production of each group balances its losses due to consumption by its predators, fishing, emigration and natural mortality. Ecosim models are dynamic simulation models which are initialized using parameters and biomasses from Ecopath. Ecosim simulates the changes in biomasses and fisheries catches over time as a result of changes in abiotic forcing (e.g. primary production, temperature) and fishing mortality or effort. Abiotic forcing can be applied on primary production, to modify predator-prey interactions (like increased search rate due to increased temperature) or egg production of species for which age-structure is modelled explicitly. Fisheries can be represented by one or multiple fleets, each of which may target one or more species or age groups.

Ecopath models will be compared using the PREBAL procedure (Link 2010), which has been implemented in R for that purpose. PREBAL is an established and standardized method to test the quality of mass-balance and compare basic ecosystem properties among Ecopath models using a set of metrics and rules.

Ecosim models can be compared regarding 1) their performance in hindcasting and the robustness of this performance to uncertainty in input parameters 2) the robustness of model forecasts to uncertainty in future trajectories of forcing.

Ecosim has a built-in function which returns a sum of squares (SS) value indicating how well the model fits to historical data time series overall (i.e. for all groups); a lower SS means a better fit, with model estimates closer to observed data. This SS value is useful when fitting the model to assess, for example, how much a particular group/parameter contributes to reducing the SS. However, it cannot be used to compare different models because: (i) SS will vary with the number of groups included in the model and the number of historical data time series loaded to fit the model, (ii) SS will vary with the length of the historical data time series used and (iii) different models may be fitted using historical data time series of different quality (i.e. noise, missing values, etc.) which will directly impact SS. To compare different models' abilities to hindcast it is more appropriate to investigate individually for each group how well the biomass/catch estimates replicate historical time series of observed biomass/catch data. To do so, Spearman's rank correlation coefficient will be calculated for each group in each model. This coefficient has a value between -1 and 1 and indicates how well the model estimates correlate with the data: 0 means no correlation, 1 means perfect correlation and -1 perfect negative correlation. Being a rank correlation it is less affected by the amplitude of the variation ('noisiness') in the data, thus providing a relatively reliable indication of whether the model captures the trends of the observed data. Each model will then be scored according to the proportion of groups achieving a correlation coefficient above a certain threshold, e.g. 0.25, 0.5 or 0.75: the higher the number of groups with high correlation the better the model hindcasting ability score. A sensitivity analysis will also be performed to assess how sensitive each model is to parameter uncertainty. To do so, the robustness of the hindcasting ability of each model will be evaluated by testing the relative changes in the score to changes in input values of biomass.

For comparing forecasts, all of the indicators described in Table 2 and their time series can be calculated using EwE models. For testing the robustness of the forecasted values of these indices to variations in future forcing, the MultiSim tool of Ecosim will be used. Using this tool, EwE models are run forced by different future realizations of the same forcing series. Then the indices of interest are computed from each model run and their resulting variability compared. The various future realizations will be generated using an established automated procedure in the R environment. This procedure generates a large number of potential future trajectories from any time series based on its statistical properties, using exponential smoothing of the time series. A common protocol implementing this method has been established for the EwE models within MareFrame.

Other models

In addition to Gadget and EwE which are the main modelling frameworks used in MareFrame with one or both used in most of the case studies, a number of other models have been applied. They include the so-called “Green model” and “Charmingly Simple Model” (see Deliverable D4.3-4.6) from the North Sea case study, a multispecies production model (Horbowy 2005) from the Baltic Sea case study, simple models used for comparison purposes in the South Western Waters Gulf of Cadiz case study.

The Green Model is a “front end model” designed to be friendly to stakeholder use. Presently it acts as a front end to the ICES WGSAM Stochastic Multispecies Model (SMS) but could be adapted to other models. Presently it is only adopted for the North Sea SMS. It works by approximating the behaviour of more complex models as a multispecies Schaefer model or a multispecies Fox model (Pope 89a, Collie et al. 2003). As such it is based upon the Jacobian Matrices of Steady State Yield or of Steady State biomass (total or SSB). In practice these matrices are calculated by performing a series of small changes in each variable of the underlying model to estimate (∂(Yield-Speies S,FleetA))/(∂Effort-FleetB) or (∂(Biomass-Species S))/(∂Effort-FleetB).

These matrices may also be calculated directly from the steady state biomasses of species at age, their partial predation mortalities at age, partial F’s at age and from the stock recruitment functions adopted (Pope 1989b). Jacobian matrices provide a powerful tool for comparing ecosystem models since they indicate the direction of change to be expected from changes in the fishery mortality drivers of the system. The theoretical form provides the additional advantage of suggesting the likely sensitivity of models to input error or parameter changes.

In relation to comparison of results for the same model, various implementations of SMS exist. The possibility of using the sensitivity of the Jacobian to model terms could also be explored. These approaches might be adopted for intra or between model comparisons.

The Charmingly Simple Model (CSM) is a length based model initially designed to investigate the relationship of size spectrum slope and other aggregate size indicators of ecosystem health to fishing pressure. Simplicity is the key to this method and many input parameters are fixed either with respect to life history invariants or to the broad results of more complex ecosystem models. However, runs with species specific fishing mortality drawn from SMS showed a surprising ability to simulate the North Sea system. Hence, developments of tuning the model to survey size data and to the moments of the catch at size data seem indicated. Because of its focus on size rather than species as such this model in its simpler versions is well adapted to considering the consequence of climate change scenarios (when the species present in region might change dramatically). In its more developed form it should be capable of most comparative metrics at a regional level.

The simple models used in the South Western Waters Gulf of Cadiz case study are models previously developed for data limited approaches. They have been already implemented in a user-friendly tool: the Data-Limited Methods Toolkit (Carruthers and Hordyk, http://www.datalimitedtoolkit.org). "The Toolkit recognizes which of the acceptable methods can be applied with the data for the stock and then applies them, generating explicit guidance for fisheries managers". The description of the methods used for Gulf of Cadiz anchovy stock according to data availability can be found in more detail at https://cran.r-project.org/web/packages/DLMtool/DLMtool.pdf, and summarized here:

SPMSY: The simple method for estimating MSY from catch (Martell and Froese, 2013).

SBT1: A model based on the management procedure used in the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) http://www.ccsbt.org/site/recent_assessment.php.

Fdem_CC: A demographic approach to calculate FMSY (McAllister et al. 2001)

CC1 and CC4: constant catch management procedures from Geromont and Butterworth (2014)

BK_CC: The estimation of potential yield and stock status from life-history parameters (Beddington and Kirkwood, 2005).

AvC: Average of historical catch

The tool uses subsets of the information available and provides a TAC given by the product of Fmsy and estimated biomass in the last years as a comparative metric. This tool can be modified to include a TAC estimation from Gadget encompassing all the methods used for the SWW Gulf of Cadiz case study in the same framework.

Atlantis

Atlantis is a whole-of-system ecosystem box model. The Atlantis model is modular and consists of various sub-models that govern the oceanographic, biological, and human components of an ecosystem. The model is 3-dimensional, spatially explicit, and consists of multiple horizontal boxes, which in turn consist of a varied number of vertical boxes based on the depth of the water column. Organisms in Atlantis can either be represented at a species level, or more commonly as functional groups. The resolution of these representations can either be at the level of gross biomass pools (in mg N / m^3 or mg N / m^2 depending on whether they are in the water columns or living on the sediment interface) or in age-structured groups (e.g., typically for vertebrates and larger bodied invertebrates like cephalopods). All the major biological processes (e.g. migration, recruitment, growth, predation) are explicitly handled. Additionally, a fisheries model can be used to parameterize fishing fleets in order to "fish" the ecosystem. Atlantis is capable of both hindcasting and forecasting. Various metrics can be obtained from Atlantis related to biomass, growth, production, landings, mortality, and predation either directly or can be easily computed. In addition, all of the indicators listed in table 2 can be readily calculated from Atlantis output.

Comparing models in practice

Model comparison is synchronized via data extraction from 'mfdb' – as a MareFrame standard to enhance ecosystem model comparison, it is recommended that competing models point to the same disaggregated data source stored in mfdb. Wherever possible this should include datasets which need to be aggregated at different levels to be used as model input data by the alternative models. Datasets recommended as part of the information stored into mfdb include:

- Commercial catch amount
- Commercial catch composition
- Survey data
- Stomach data

Identification of common scales of comparison – model resolution may differ considerably among the ecosystem models used in MareFrame. To make the application of each model as efficient as possible it is advised that models designs exploit and best reflect the strengths of each modelling framework rather than attempting to force different models into structurally similar forms. The lack of a single form may mean model comparison is not as straightforward, however, meaningful comparison of indicators between models may still be possible once the indicators are computed and then aggregated over a common resolution. Identification of a common resolution should ensure that comparison is carried out on comparable:

- spatial units - of relevance for spatially disaggregated models or when models differ in their respective study areas
- temporal units – as some models may implement seasonality and use a timestep smaller than 1-year
- biological units - depending on the indicator, different models may provide information on parts of a population (i.e., certain age/length groups, life stages such as juvenile and mature), on entire individual populations, on meta-populations (i.e., in the case of multiple stocks) or functional groups (i.e. group of species)
- other units – for instance in the case of multi-fleet models, catches have to be aggregated over so called 'pseudo-fleets' if required before comparison.

Link case study primary objectives to the dynamics of specific model components – to enhance the model comparison, specific model components which better link to the primary objectives of the case study should be identified. Priority should be given to comparison of metrics on guilds, functional groups and model components which are most effective on the ecosystem dynamics and their assessment, or because of their direct link with the management issue. Biological groups should be identified according to the following three categories: (1) groups with a fast turnover which may represent "early warning" indicators, (2) groups targeted by fisheries in which there is an economic interest, (3) charismatic or vulnerable groups which tend to include species with a slow growth and long life-cycles that may integrate across a range of long-term impacts, and are often of conservation concern. Key functional groups (e.g., forage species and top predators), trophic groups (e.g., piscivores and planktivores) and habitat-associated groups (e.g., coastal and off-shore) should be identified according to the specificities of the case study.

Comparison matrix – to facilitate comparison of different models over several management scenarios, model performances can be organised in tables where models and scenarios represent rows and columns respectively, and a specific metric for comparison is tabulated together with its upper and lower boundaries (i.e., median and 95% CI).

Graphical representations – these include so-called 'Zeh-plots', traffic plots and plots of single indicators for direct comparison across models

Management Strategy Evaluation comparison framework: The advantage of having a reference population created with operating models, and considered as the "truth", is that it allows testing of model performance as the difference between estimation (associated to models) and calculation (associated to reference modelled population) of indicators. In order to include variability, the intrinsic variation of estimated and calculated catches time series can be used to give a rank of possible values of the specific metric chosen for comparison.

Evaluation and feedback regarding the application of models in case studies

As the project has been developed on a co-creation with stakeholders basis, one of the axes of evaluation incorporates stakeholders' view. The DST has been the interface between case study models and *stakeholders*. Stakeholders have tested the models developed and have suggested new modifications in one or two rounds of meetings (according to the case study). Their feedback is a usefulness indicator of the models developed and how each case study model have include stakeholders suggestions or concerns, evaluate the progress of modelling process as could be seen in Figure 4.

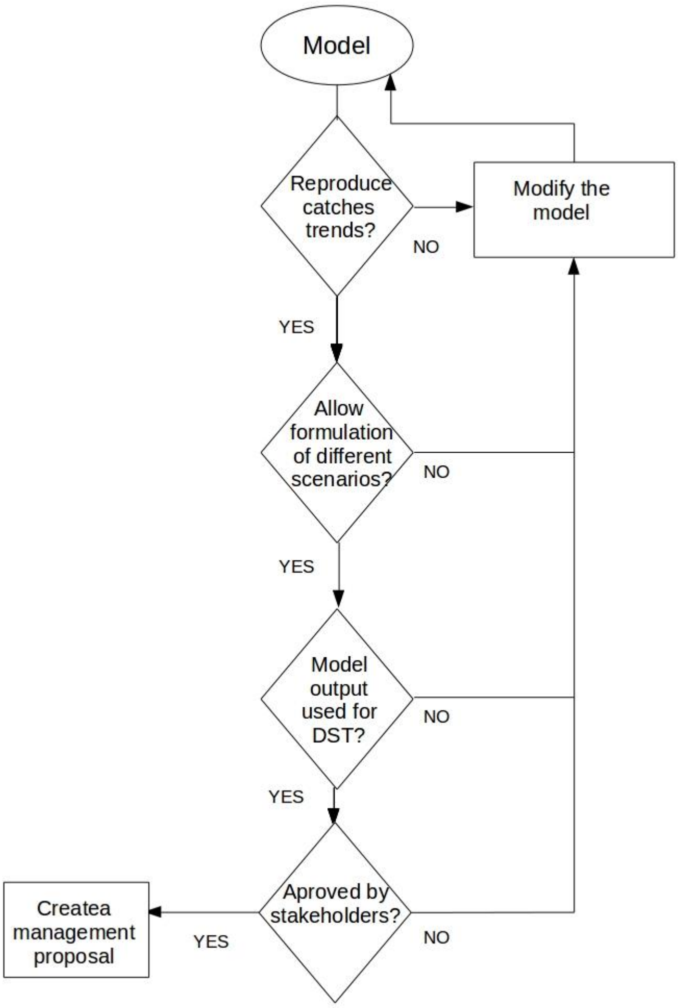

Figure 4: Flux diagram of evaluation of the application of models in case studies.

Baltic Sea case study

Current fisheries management of the three most commercially important stocks in the offshore Baltic Sea is currently based on TACs which are specified annually. EAFM in the region likely will use more long-term management plans which require a vision and strategy to achieve a desired future state of the ecosystem. Such a strategy needs to be grounded in scientific knowledge on what are the possible future states of the ecosystem, how far they are from desired states and what are the crucial trade-offs between management outcomes. To address that need, the modelling focus in the case study has been on simulating mid-and long-term consequences of management scenarios. Although the modelling approaches used in the case study are generally thought to be useful on different scales, their comparative strengths and weaknesses have rarely been tested systematically to prove that. Therefore, the strategy of the case study is to simulate management scenarios as consistently as possible with each three model and evaluate their usefulness for answering various questions based on that exercise. Scenario simulations are not finalized with all three frameworks, thus, the present comparison is restricted to a comparison of the scope of models and their hindcast abilities.

We used three types of modelling approaches in the case study: Ecopath with Ecosim (EwE), Gadget (both single-species and multi-species implementations) and a Multispecies Stock-Production Model (MSPM). Table 3 summarises the application stage of each model within the case study. In addition to the general model evaluation scheme common to all case studies (refer to flowchart figure), there are a few case-study-specific steps for model application. For the Baltic case study these include the implementation of climate, nutrient and seal population growth scenarios, as far as model structure allows, to estimate environmental uncertainty of projections. Additionally, projected time series (2013-2030) of indicators have been binned and formatted for visualization and use in the case study Decision Support Tool (DST, refer to ST6.3.1).

The EwE model is fully implemented (with the exception of the multispecies F_{MSY} scenario) and linked to the DST. Preliminary model outputs were used in the workshop to introduce stakeholders to the first prototype of the decision support framework and DST (7th October 2016, MS18). At the time of writing of this report, management but not environmental scenarios have been run by the MSPM and results have not been linked to the DST yet, while the Gadget model has been parameterized but forecasts are limited to the BAU management scenario.

Table 3 Overview of application stage of each model in the case study

Application stage	EwE	GADGET	MSPM
Model design completed	✔	✔	✔
Input files for fishing mortality scenarios prepared	✔	✘	✔
Input files for environmental scenarios prepared	✔	✔	✘
Model is fully parametrized	✔	✔	✔
Model creates reasonable/expected results	✔	✔	✔
Output linked to ecological and economic indicators	✔	✔	✔
All management scenarios modelled in satisfactory way	✘	✘	✘
All cross-combinations of management and environmental scenarios modelled	✘	✘	✘
Outputs prepared for input into DST	✔	✘	✘
Outputs presented to stakeholders	✔	✘	✘

The most important considerations for an EAFM in the Baltic were defined in cooperation with stakeholders. As can be seen from table 4, there is no single model which is able to address all considerations, but the three models together achieve this. However, it has to be noted that more in depth analysis of socio-economic data (and perhaps additional economic modelling) would be necessary than what is currently possible in the project to address the socio-economic benefits derived from fisheries in a more satisfactory way.

All models provide information about the biomasses of all three species of interest (cod, herring, sprat) in the Baltic, with EwE and MSPM providing limited information about internal structure (both simulating only a few stanzas per species) and Gadget representing internal size-structure in detail. On the other hand, EwE and MSPM both incorporate the dependency of predator (cod) biomass growth on prey availability and the effects of different environmental drivers on the stocks, which are not yet represented in our Gadget model. In EwE there is a link from cod reproductive volume to cod reproduction, temperature to sprat and herring reproduction and zooplankton feeding, and from hypoxic area to benthos feeding. In MSPM hypoxic area is directly linked to cod growth and salinity to herring growth. In all three models, the representation of fisheries in the scenarios is the same, although EwE and Gadget to some extent have broader capabilities in that respect. Economic and social indicators are calculated in the same way in all three models. Seal effects on small-scale fisheries are considered in the same way in all three models as well: seal biomass influences profits calculated for gillnet fishery.

Table 4 Summary of relevant ecosystem components for magament and whether they are modelled within the modelling approaches.

Management considerations	EwE	GADGET	MSPM
spawning biomass	✔	✔	✔
size structure	✖	✔	✖
low growth and poor condition of cod	✖	✖	✔
trophic interactions among cod, sprat, herring stocks in the Central Baltic	✔	✔	✔
non-commercial key food web components as in MSFD (fast-turnover groups, charismatic species)	✔	✖	✖
major environmental drivers influencing the dynamics of the harvested populations	✔	✔*	✔
hydrography/eutrophication linked to reproductive conditions	✔	✔*	✖
social and economic benefits derived from fisheries	✔	✔	✔
competition between small-scale fishery and seals	✔	✖	✖

* limited to the way how recruitment is generated in the forecasts

Comparison of model hindcasts

The three models were independently parameterised from each other, but used similar input data, surveys, catches and assessments to some degree (see D5.3 for details). They are all spatially non-explicit and have yearly (EwE and MSPM) or quarterly (Gadget) time steps. To compare their historical projections with respect to biomass, we calculated the mean biomass from the three models ('ensemble mean') per stanza, as well as for demersal to pelagic ratio and investigated model deviations (Figure 1,2). The models generally agreed in the trend, with larger uncertainties for juvenile groups. Gadget hindcasts of juvenile sprat and herring (and to some extent adult herring) are generally lower than those of the other two models (Figure 1). This is probably one reason why Gadget hindcasted a higher D/P than MSPM, with EwE values being between the two (Figure 2). There are no systematic large deviations between the model output and historical catch data. There are a few overestimates of cod for the later years by Gadget, mostly related to overestimations in the active fleet catches, where cod catches were estimated based on effort data, while sprat catches are somewhat overestimated by EwE (Figure 3). The described model differences will carry over into the forecasts and will contribute to estimate the forecast uncertainty derived from different modelling

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

approaches. Moreover, these analyses are the first step for understanding how model differences can be related to differences in parametrisation, input data or model processes, which is crucial for a better understanding and interpretation of scenario forecast results.

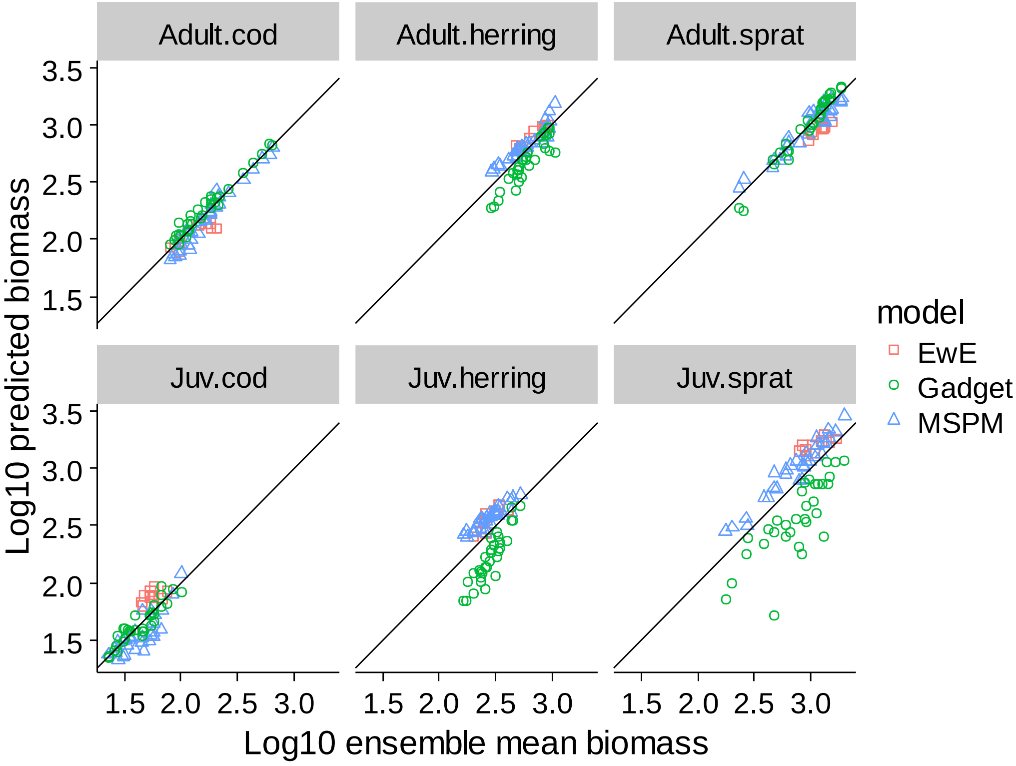

Figure 1. Log 10 predicted biomasses by three models compared to the ensemble mean. Dots represent individual years. MSPM and Gadget hindcast is compared 1981-2013, EwE hindcasts are only done for the period 2004-2013.

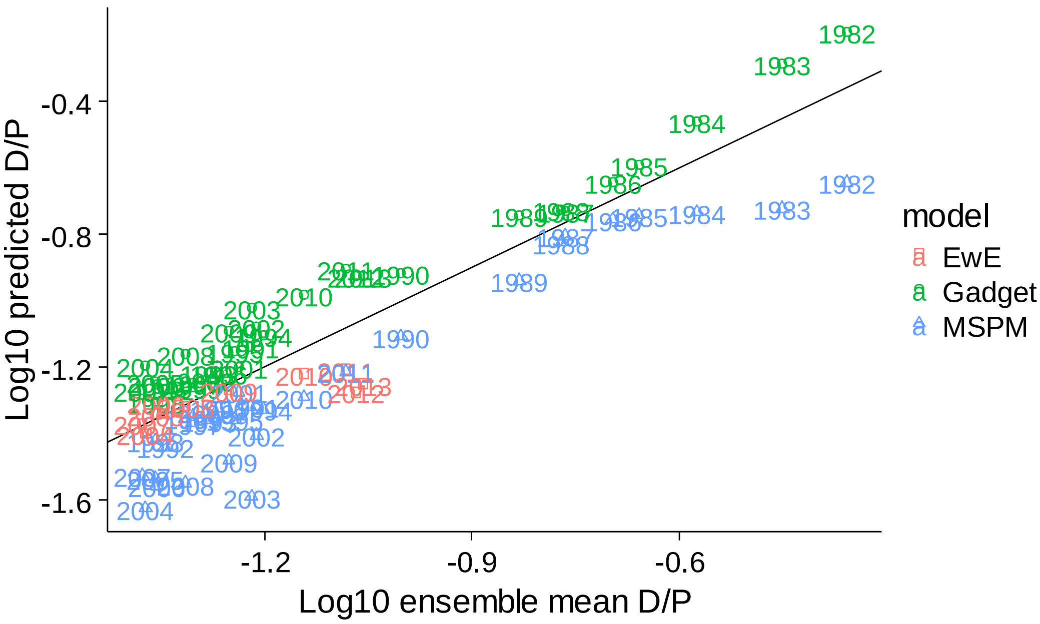

Figure 2 Log 10 predicted biomasses by three models compared to the ensemble mean. Dots are labeled with the years they represent. MSPM and Gadget hindcast is compared 1981-2013, EwE hindcasts are only done for the period 2004-2013.

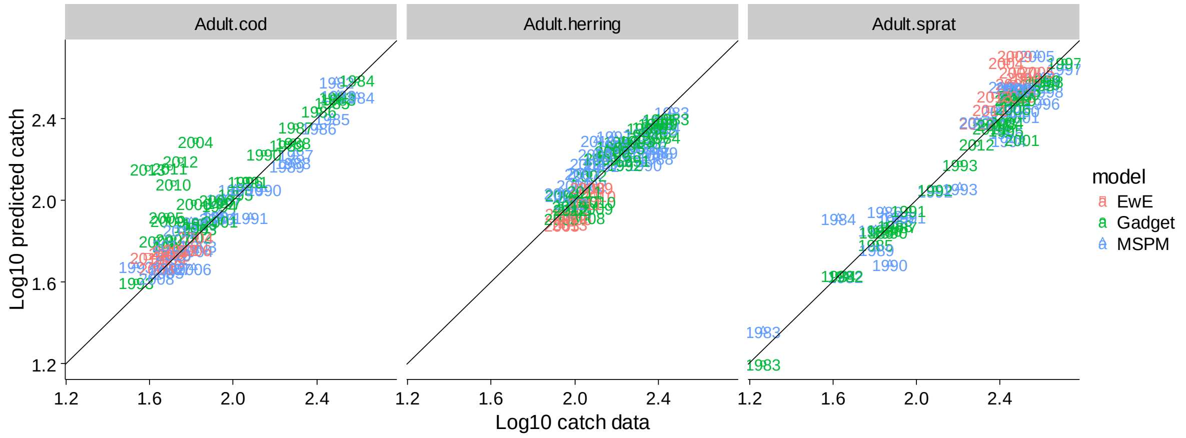

Figure 3 Log 10 predicted catches by three models compared to data. Dots are labeled with the years they represent. MSPM and Gadget hindcast is compared 1981-2013, EwE hindcasts are only done for the period 2004-2013.

Comparison of model uncertainty

Although EwE is the most complex model applied in the case study, and therefore has the largest number of free parameters and the most subjective model-building process, it has fairly advanced built-in routines to test sensitivity to many of these parameters, and further development of such functionality is expected. The routine Monte Carlo simulation was used to evalute the sensitivity of Ecosim hindcasts to Ecopath input parameters (see D5.3), with larger effects on juvenile cod biomass, adult flounder biomass and adult cod and flounder catches when compared to other fish groups. Similar analyses on scenario outcomes could be conducted but have not yet been done. Monte Carlo analysis doesn't replace personal expertise of the modeller, which Ecopath more heavily relies on than the other two approaches, given its larger scope. Ecopath parameters and partly Ecosim settings of our model were reviewed by experts in the ICES Working Group on Multispecies Assessment Methods, which is probably the most reliable method in case of Ecosim models to avoid personal biases or subjective parametrization as far as possible. Full sensitivity analysis to all Ecosim settings in all scenarios is not feasible, however, a systematic exploration of a few selected parameter combinations are possible. These combinations could include those parameters that infuence hindcast fit the most (especially those of adult cod). Currently, sensitivity analysis on all multispecies Gadget model parameters is also limited by the elevated number of parameters, elevated computation time, high uncertainty in specific processes (ie, consumption), high noise in some crucial datasets (ie, stomach data), lack of predefined implemented routines. The MSPM framework appears more mature in this respect as shown by Horbowy (1996).

Once input parameters are chosen, model fitting is an objective process in all three modelling frameworks. Data weighting is another crucial aspect of model fitting which is dealt differently by the different models. At the moment objective weighting and sensitivity on the weighting are limited in our EwE implementation by the lack of automated routine available. Gadget achieve objective weighting via an iterative approach (Stefansson 2003) but at the expenses of high computation time which limits both the number of configurations which can be tested and any realistic implementation of sensitivity analysis. As indicated above, in theory all models are well suited to simulate and quantify effects of environmental uncertainty on model forecasts (see D5.3), but in the case of MSPM this is dependent on an R implementation of the model, while for Gadget it is still cumbersome as routines in Rgadget are not consolidated yet.

North Sea case study

The North Sea Case Study is potentially the most complex of all of the MAREFRAME case studies, both in number of species included (12+) and number of fleets (Circa 150 defined in terms of Country*Gear type combinations). However fortunately, the North Sea is blessed with several pre-existing Multispecies Models (see ICES 2014) that can form inputs to MAREFRAME work in addition to those models being developed under MAREFRAME. Consequently, the approach to date has been to develop a “front end” model called the Green Model following the approaches of Pope, (1989) and Collie et al (2003). This is designed to be user friendly for stakeholders and to provide them with an understanding of the trade-offs and also to provide all necessary outputs for the DSF. To be usable by stakeholders it must be

1. Readily transportable to stakeholder computers - so is written in EXCEL,
2. It must be easy to comprehend – so it is driven by sliders, and has a car style dash board of results.
3. It must respond very rapidly to queries - so that it is based upon a quadratic approximation fitted to other more complex multispecies models.

Despite these constraints it is able to deal with

1. the response of a complex models of the multispecies system to changes in fishing mortality on any of the 12 species,
2. The constraints imposed upon achievable fishing mortality by the fleet structure
3. Provide Social and Economic Trade-offs in addition to the biological outputs of catch, SSB and where possible discards.
4. As many of the GES indicators as possible.
5. It is also capable of making constrained optimisations of measures of interest to the DSF in a matter of a few minutes.

The Green Model‘s underlying multispecies model is a multispecies Schaefer model (a quadratic function of the fishing mortalities). This might in principle be fitted to historical data but given the potentially large numbers of parameters to fit, N*(N+1) where N is the number of stocks (N=12, Potential parameters=156 in this case), this is not a very attractive option. The alternative is to fit the parameters to a current estimate of steady state yield (typically the long term steady state at status quo F) and to the Jacobian matrix about this state of species yield or species SSB or species discards with respect to the fishing mortality on each other species derived from the results of other models that draw on biological data. That is to derive the value of Y’(i) and of ∂Y’(i)/∂F(j) for i and j =1: N where Y’(i) is the steady state yield of species i and F(j) the fishing mortality rate on species j. One virtue of this approach is that providing that any multispecies model can be run to estimate these, (i.e. the steady state solution and the consequence to the steady state landings or SSB or discards of increasing fishing mortality rate on each species in turn by say 10%) its interpretation of the long term behaviour of the North Sea can be easily inserted into the Green model and the comparisons required for Delivery 7.2 easily generated. This makes comparisons between multispecies models of the consequences of the biological interactions easy and means they are all compared using the same fleet constraints and the same social and economic model.

So far the Multispecies Models available or being considered are as follows.

1. SMS- this is the standard North Sea model (Lewy and Vinther, 2004)based mostly upon species catch at age data tuning data and stomach content data mostly collected in the two years of the stomach (1981 and 1991).
2. Le Mans - this is an Ensemble model of the North Sea (Thorpe et al, 2014, 2015) that tries to find alternative parameterizations that satisfy the historical time series. The current comparison is based upon only 1 plausible run of this but Robert Thorpe its author is collaborating to provide an ensemble of his best set of models.
3. We have a Multispecies Schaefer fitted by taking the approximation to SMS as the starting point and allowing those interaction terms that are not included in the SMS stomach data sets to be modified by regression on historical catch and mortality time series. This was presented as a poster at ICES ASC 2016 and will be included as a comparison.
4. A full MS upgrade to the CSM (Pope et al 2006) is nearing completion. Earlier models are promising and already provide approximations for size GES indices such as the Large Fish Index. They also provide a basis for examining the likely consequences of climate change since they are not tied to existing inventories of species but instead define species by L_{∞} groups. The extended model will also allow novel trophic level data to be included.
5. Experience with the existing North Sea multispecies models suggests that the largest multispecies effects are on the recruitment levels for prey species. This suggests that fitting a delay difference model of the system may be a better way to fit historical data than directly fitting a Schaefer model. This should not be a lot of work and should translate easily into a Schaefer model for long term predictions.
6. EwE (Mackinson & Daskalov, 2007) - results from an updated version of this (Mackinon pers comm) version of this are available but it will not fit into the full Green Model because it has its own fleet structure, and so cannot provide a Jacobian Matrix by fishing mortality rate (only by fleet). It does however provide a current status quo and can provide 25% up and down on all Fs solutions.
7. We promised a Gadget model for the North Sea but this is looking problematic. Advice received (D. Howell pers. comm) is that this is not a good idea for the North Sea (basically just too big and complex). If this is done it will be as Gadget-lite using single species biomass as tuning series.

Of these 1 and 2 are shown in the existing comparison of long term behaviour in Table 2 below. Models 3 and 6 could be fairly readily added.

All of these models have potential uses.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

Table 1: Overview of application stage of each model in the case study

Application stage	Green	SMS	Le Mans	MS Sch'fer	CSM	EwE	Delay Diff.	GADGET
Model design completed	✓	✓	✓	✓	✓	✓	✗	✗
Input files for fishing mortality scenarios prepared	✓	✓	✓	✓	✗	✓	✗	✗
Input files for environmental scenarios prepared	✓	via Green Mod. ✓	v G M ✓	v G M ✓	✓	✗	✗	✗
Model is fully parametrized	✓	✓	✓	✓	✓ only Loo form	✓	✗	✗
Model creates reasonable/expected results	✓	✓	✓	✓	✓	✓	✗	✗
Output linked to ecological and economic indicators	✓	vGM ✓	vGM ✓	vGM ✓	✓	✗ limited	✗	✗
All management scenarios modelled in satisfactory way	✓	vGM ✓	✗	✗	✗	Not poss ✗	✗	✗
All cross-combinations of management and environmental scenarios modelled	✓	vGM ✓	✗	✗	✗	N. p. ✗	✗	✗
Outputs prepared for input into DST	✓	vGM ✓	✗	✗	✗	N.p. ✗	✗	✗
Outputs presented to stakeholders	✓	vGM ✓	✗	✗	✗	✗ N.p.	✗	✗

Past experience with multispecies models in the North Sea suggests they differ significantly from single species models in terms of long term steady state yield but rather little in terms of short term predictions. This suggests that they should be treated as strategic models. This is somewhat at variance with experience in more wasp-waisted systems than the North Sea e.g. the cod and capelin in the Barents Sea where predation can influence short term management decisions.

Comparisons between Multispecies models so far suggest that the Ensemble model tends to choose parameterizations with less multispecies interactions than are found in SMS. Comparisons made for the ICES ASC 2016 poster between SMS and the Schaefer model suggest quite large differences in steady state yields as a result of adding/modifying the interaction terms.

It seems that differences between models are likely to be the dominant source of variance that we need to consider. Hence, we will need to consider carefully how best to advise from contradictory models. One suggestion is Nash type (mini max) (Nash 1951) solutions may be more stable than full maximizations of particular objectives.

Table 2: Comparisons of long term results made so far.

	SMS based GREEN MODEL	THORPE MODEL run 10 Based GREEN MODEL
Indicators		
Biological Reference Point based indicators		
Number stocks		
	Above Blim?	Above Blim?
F at 125%	12	10
F at 100%	11	10
F at 75%	11	10
Biomass based indicators	Species	Species
Biomass	ALL except MAC and NEP	ALL except MAC and NEP
F at 125%	4003695	5997421
F at 100%	4264616	6951105
F at 75%	4525538	8104821
Abundance	N/A	N/A
	Species	Species
Catch	ALL except MAC and NEP	ALL except MAC and NEP
F at 125%	1627922	2154134
F at 100%	1403190	2062300
F at 75%	1128031	1800969
Fishing revenue	GEARS	GEARS
	all	all
F at 125%	€ 1,799,544,189	€1,997,660,939
F at 100%	€ 1,599,569,510	€1,836,080,848
F at 75%	€ 1,328,035,003	€1,563,931,960
Fishing mortality		
	Average all bar NEP	Average all bar NEP
F at 125%	0.43	0.46
F at 100%	0.35	0.37

F at 75%	0.26	0.28
Ratios indicators		
Pelagic:Demersal	Pelagic:Demersal	Pelagic:Demersal
F at 125%	4.31	2.40
F at 100%	3.72	2.19
F at 75%	3.32	2.06
NB. exploitable biomass of (HER. +.1*NEA MAC +NOP+SAN+SPR) /exploitable biomass(COD+HAD+PLE+POK+SOL+WHG)		
TopPredators:ForageFish	NA	NA
Planktivores:Piscivores	Planktivores:Piscivores	Planktivores:Piscivores
F at 125%	10.83	5.04
F at 100%	9.47	4.72
F at 75%	8.55	4.52
NB this is exploitable biomass of (HER. +.1*NEA MAC +NOP+SAN+SPR) /exploitable biomass(COD+POK+WHG)		
Catch:Biomass	Catch:Biomass	Catch:Biomass
F at 125%	0.36	0.35
F at 100%	0.29	0.28
F at 75%	0.21	0.21
Biodiversity indicators		
Shannon's diversity index based on landings		
F at 125%	9.26	6.51
F at 100%	9.62	6.83
F at 75%	9.79	7.06
NB Based upon the catch of the 12 species		
Number of species with landings > minimum level		12 for all scenarios
Trophic indicators		
Mean Trophic Level (MTL)	NA	NA
Marine Trophic Index (MTI)	NA	NA
Essentially this is meaningless for the North Sea where diversity varies with size		
Size-based indicators	LFI	LFI
Large Fish Indicator F=125%	52%	54%
Large Fish Indicator F=100%	59%	61%
Large Fish Indicator F=75%	65%	66%
Economic indicators	all fleets	all fleets

Revenue		
F at 125%	€ 1,799,544,189	€1,997,660,939
F at 100%	€ 1,599,569,510	€1,836,080,848
F at 75%	€ 1,328,035,003	€1,563,931,960
Profit		
F at 125%	€ 409,084,206	€607,200,956
F at 100%	€ 485,533,955	€722,045,293
F at 75%	€ 490,423,874	€726,320,831
Social indicators Employment		
	at sea employment costs	at sea employment costs
F at 125%	€ 477,260,442	€477,260,442
F at 100%	€ 382,355,567	€382,355,567
F at 75%	€ 287,450,691	€287,450,691
	Process Labour Fish Meal FTE	Process Labour Fish Meal FTE
F at 125%	1157	1245
F at 100%	1030	1186
F at 75%	859	1040
	Process Labour Rest FTE	Process Labour Rest FTE
F at 125%	11066	12378
F at 100%	9832	11276
F at 75%	8155	9534

Comparisons of interactions using the Jacobian Matrix

An interesting way of comparing multispecies models springs from the approximate fits of the Multispecies Schaefer models to the results from steady state solutions of more complex models. The Schaefer model describes steady state yield in terms of fishing mortality by N equations defined by;-

$Y^1(i) = A(i)F'(i) + F'(i) * \Sigma_{\text{all } j} (B(i, j)*F'(j))$.   Equation NS1

Where $Y^1(i)$ is the steady state yield of species i, when fishing mortality rate on all species are set to status quo, i.e. when all F'(i) =1 where F'(i) Fishing mortality rate of species i written as the proportion of its status quo F. Ai and B(i, j) are the constant terms of the quadratic equation and both i and j=1:N where N is the number of species included in the ecosystem model.

This approximation is constructed from outputs from more complex models of yield (and also for discards (where known) and SSB) at the future steady state at status quo fishing, together with the equivalent long term steady states to be expected with a 10% increase in each individual species fishing mortality above status quo. This is the minimum information needed to solve parameters. Since if we know the $Y^1(i)$ and the $Y^{1.1(j)}(i)$ from the more complex equation where $Y^{1.1(j)}(i)$ is the steady state yield when F'(j) is set to 1.1 and all other F'(i) = 1 where j ≠ i.

Then where j ≠ i

$(Y^{1.1(j)}(i) - Y^1(i))/0.1 = B(I, j)$ Equation NS 2

Where i=j

$(Y^{1.1(i)}(i) - Y^1(i)) = A(i)*.1 + B(i,i)*.21$ Equation NS 3

We may then substitute the B(i ,j) results of equations 2 back in to equation NS1 to get

$Y^1(i) = A(i) + B(I, i) + \Sigma_{\text{all } j\neq i} (B(i, j)$ Equation NS 4

and we may solve equations 3 and 4 for A(i) and B(i, i)

Note that equation 4 indicates that the sum of A(i) + $*\Sigma_{\text{all } j} (B(i, j) = Y^1(i)$

Hence dividing all A(i) and all B(I,j) by $Y^1(i)$ standardizes them to a sum of 1. This is convenient for comparing the sizes of interactions of different models.

For comparative purposes it is more revealing to work with the Jacobian Matrix itself. When the a(i) and B(I, j) have been calculated N*N the Jacobian Matrix $J=[\partial Y^1(i)/\partial F'(j)]$

This may be written as J=B+diag(A)

where B is the N*N matrix of the B and diag(A) is an N*N matrix with A(i) as the main diagonal and zero elsewhere. Again it is convenient to standardize the Jacobian values by dividing by $Y^1(i)$.

Figure NS1 below shows radar plots for the $\partial Y^1(i)/\partial F'(j)$ of each species derived from the SMS run used to parameterize the GREEN model and an alternative parameterisation based upon a single realization of the Thorpe *et al* (2015, 2016)Le Mans model.

Note that if all $\partial Y^1(i)/\partial F'(j)$ were zero when i ≠ j then we would just have a series of single species models. This seems very nearly to be the case for plaice(PLE), saithe(POK), sole(SOL) for the Green Model and for most species for the Thorpe results. It is far from the case for cod(COD) and whiting(WHG) for the Green model. That the Green Model Results are more reactive than the Thorpe results is apparent on this plot.

What should we make of this? First it would seem how interactions are estimated has a large impact upon their size. It is perhaps interesting that haddock, herring, Norway pout and whiting all have noticeable interactions with pok (saithe). The saithe was the species where the stomach content data collected in 1981 and in 1991 gave notably different estimates of suitability for MSVPA. (Rice et al 1991 and ICES 1992, ICES,1994) and may give rise to particularly large uncertainties.

It will be interesting to add other specimen models to this Figure!

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

Figure 4: Comparison of standardized interaction terms for each species derived from the le Mans and from the Green model. The i in the labels should be understood to be the species given in the legend.

Icelandic case study

Fisheries in Icelandic waters are generally controlled through a ITQ system, where TACs are set for individual species. These TACs are usually based on a harvest strategy which includes an assessment using Gadget as a tool for stock estimation. Although there are a few exceptions, the annual assessments do not take into account multispecies issues, nor do allocations normally account for technical interactions. Issues other than direct stock estimation are usually not addressed through the assessment process. Thus, there are other procedures for considering the effects of bottom trawl on the benthos or the effects of mesh sizes etc.

For this case study three basic approaches have been considered:

- A Gadget multifleet model
- An EwE model
- An Atlantis whole-of-system model

As outlined above, the first issue is "the capacity to support an ecosystem approach to fisheries management" and this is well summarised in stating that they should "provide qualitative and quantitative estimates of the expected benefits, costs, and risks associated with alternative management actions". The following starts by reporting some of the overall conclusions from the modelling effort, with an emphasis on which models can be used for which purpose in the Icelandic context.

The Gadget implementation is a formal statistical approach, describing each species separately based on fits to data, and then linking the species together in a multifleet fishery. Thus this model at present takes into account technical interactions, which are a major issue in any TAC system. The Gadget methodology is the baseline methodology for Icelandic waters for a reason: The methodology satisfies the basic criteria for an assessment, being based on a formal statistical approach, which can be extended to include new processes as needed. The approach is as objective as any statistical method in that estimation is fully objective but as usual, the choice of which processes to include or how is quite subjective.

The EwE approach is the direct opposite of the Gadget approach. An EwE model starts with a balanced Ecopath model and this is also where the fundamental problem with EwE starts. Traditionally the Ecopath user picks up information from available sources to balance the starting model. Also traditionally, these will be biomass values from existing assessments and/or fishing mortality values as well as any other data needed. In all tests for the marine ecosystem around Iceland this approach has turned out to be completely subjective and can apparently lead to totally different models when undertaken by different individuals. Another problem is that the Ecopath model can be asked to estimate certain unknowns assuming the values of other things. First of all, this "estimation" ignores data uncertainty. Secondly one has a choice of which unknowns to estimate and which to enter. Thirdly, the "data" used are not "data" in the sense used in science: These are not measurements but instead are either outputs from models or poorly justified values, taken from international databases such as FishBase, where values are sometimes assumed to hold across multiple stocks in several ecosystems. This runs contrary to all scientific methods where statistical approaches are used to fit to scientific measurements. Finally, the approach of using for “data” biomass outputs from single species models directly contradicts the intent to provide multispecies assessments.

These Ecopath problems can be alleviated somewhat by the user carefully going through each part of the model and making sure it "makes sense", i.e. verifying that the "estimated" values are reasonably close to whatever should be expected. Unfortunately this is (a) subjective and (b) again relies on comparisons to single species assements (i.e. absolute stock sizes).

Formally fitting the Ecopath equations to actual measurements would seem to be the way forward, to objectivity. Similarly, fitting an Ecosim model to data will also provide more objectivity but when compared to a Gadget model fit, the entire process is a bit like comparing an old-style fit-by-eye to todays statistical model fits.

The approach with the Atlantis model has been to include not just all the main commercial species, but also the entire system, from hydrography up through economics of the fishery and social implications. It has become clear through this development that Atlantis can reasonably well replicate existing time series of information on stock trends. Thus Atlantis has a large number of internal parameters which can be adjusted so that the model behaves in accordance with expectations. Like EwE Atlantis has no method of estimation and this brings a fair amount of subjectivity into the process of defining the model. However, since the starting model is dynamic, it is certainly possible to mimick reasonably existing data trends, e.g. in survey indices and the like. Needless to say, there is no particular reason why absolute abundance values should be the same as in a traditional stock assessment, but overall trends are expected to be similar, since both should mimic survey data.

It is clear from the above that both Atlantis and EwE have problems related to subjectivity. When the Atlantis model was set up, it was "tuned" to existing stock trends. Ecopath alone has to be "tuned" to absolute stock size from a single-species assessment and this is completely illogical. Similar issues arise when initializing the first year of an Atlantis run, but normal procedure is to follow this with a comparison to data trends and the same can be done using Ecosim.

Conclusions

This entire experience leads the Icelandic partners to conclude that the Ecopath model, in its present state, is inadequate for any purpose related to management. It is, however, useful as a comparative exercise which will be continued in an effort to investigate whether the process can be made somewhat less arbitrary, e.g. through other implementations of the Ecopath model (in R) or through data-fitting exercises using Ecosim.

Since some of the same problems haunt Atlantis, it is also inappropriate for short-term assessments. This is well-known and not novel to this case-study. However, this case study has shown that Atlantis can replicate known aspects of the marine ecosystem around Iceland. In addition, Atlantis can be used to generate data for input into other models. Since Atlantis includes the entire ecosystem, from hydrography up through social impacts, it becomes the perfect Operating Model. Atlantis is also the obvious choice to answer many what-if questions, at least in a qualitative manner. Overall, therefore, Atlantis is the preferred model for generic questions on ecosystem responses, as well as representing a theoretical "truth" for comparing other models.

Finally, Gadget seems to be the only feasible modelling approach for tactical advice: This is the only approach which has objectivity and statistical estimation at its core, fits to actual data and has been developed to include the main system processes needed in assessments. The existing Gadget model for groundfish does not include species interactions, but this is not believe to be a major issue for that complex, compared to the technical interactions which are being modelled. On the other hand, biological interactions become more important once some of the pelagic species are included, which is a natural next step. Notably, the Gadget modelling environment can accommodate this approach.

Table 3 Overview of application stage of each model in the case study

Application stage	Atlantis	GADGET	EwE
Model design completed	✔	✔	✔
Input files for environmental data prepared	✔	✔	✔
Model is fully parametrized	✔	✔	✔
Model creates reasonable/expected results	✔	✔	✔
Output linked to ecological and economic indicators	✔	✔	✘
All management scenarios modelled in satisfactory way	✘	✔	✘
Outputs prepared for input into DST	✘	✔	✘
Outputs presented to stakeholders	✘	✔	✘

West of Scotland case study

The West of Scotland Ecosystem comprises the shelf area west of Scotland (ICES subarea VIa) and supports several valuable fisheries: (i) a demersal mixed fishery targeting mainly cod, haddock, whiting, European hake, saithe and monkfish, (ii)a shellfish fishery targeting the Norway lobster and (iii) a pelagic fishery targeting mainly Atlantic mackerel, horse mackerel, herring and blue whiting. These fisheries are currently managed through TACs and quotas set each year individually for each stock without multispecies considerations. Additional measures such as effort and gear restrictions and closed areas are also in place (for full CS description see D 5.1). The West of Scotland fisheries currently face several management issues. Firstly, the stocks of cod and whiting are currently depleted well beyond precautionary levels. Secondly the population of grey seals has been increasing over the past 2 decades and has been linked to an increase of predation mortality on cod which could jeopardise effort to recover the stock Cook et al., 2015). In addition, the presence of 2 depleted stocks in a mixed fishery is likely to result in choke species which will jeopardise the productivity fishery when the landings obligation comes into place in 2019. Under the MSFD, GES must be achieved by 2020. This includes bringing all exploited stocks above precautionary levels. While not all descriptors can be assessed in the a fisheries context, an ecosystem approach allowing for multispecies consideration and ecosystem indicators must be employed to identify the best management alternatives. For this case study we use to very different modelling approaches: Ecopath with Ecosim and Gadget.
Ecopath with Ecosim (EwE) is an end-to-end foodweb model representing species as functional groups (biomass pools) and fisheries (for full description see D 4.1). Ecopath is a mass-balanced model of an ecosystem where biomass production of each group balances its losses due to consumption by its predators, fishing, emigration and natural mortality. Fisheries can be represented by one or multiple fleets targeting different groups. Ecosim is a dynamic simulation models which uses parameters and biomasses from Ecopath to simulate the changes in biomasses and fisheries catches over time as a

result of fishing mortality or effort. Ecopath requires numerous parameter values to be entered, many of which are estimated based on best available data when no peer reviewed information is available for the species and area considered (e.g. diet composition). It is therefore up to the user to make sure that the model is ecologically sound. However, as the model is now used widely in marine ecology a pre-balance (PREBAL) has been developed as a standardized method which consists in a series of diagnostics and helps ensuring that the model is ecologically sound (Link, 2010). These PREBAL diagnostics were successfully applied to the West of Scotland Ecopath. Ecosim on the other hand is fitted using a statistical fitting procedure which minimises the sum of squares between the model estimates and the time series of historical data for both biomass and catches. When fitting Ecosim, catches are considered on absolute scale (i.e. the model aims at replicating the historical catch values) while biomasses are considered on relative scale (i.e. the model aims at replicating the historical trend in biomass rather than replicating the exact values). As a result, biomass time series from survey data were used to fit Ecosim for the West of Scotland in order to capture the trend shown from empirical data rather than from assessment model estimates (the exception to this were cod haddock and whiting for which data was needed for multiple stanzas). Following the fitting procedure, the West of Scotland Ecosim successfully replicates catches and biomass trends (Fig. 1). An Ecospace module which adds a 2D spatial dimension to the model is also available but is yet to be developed for the West of Scotland.

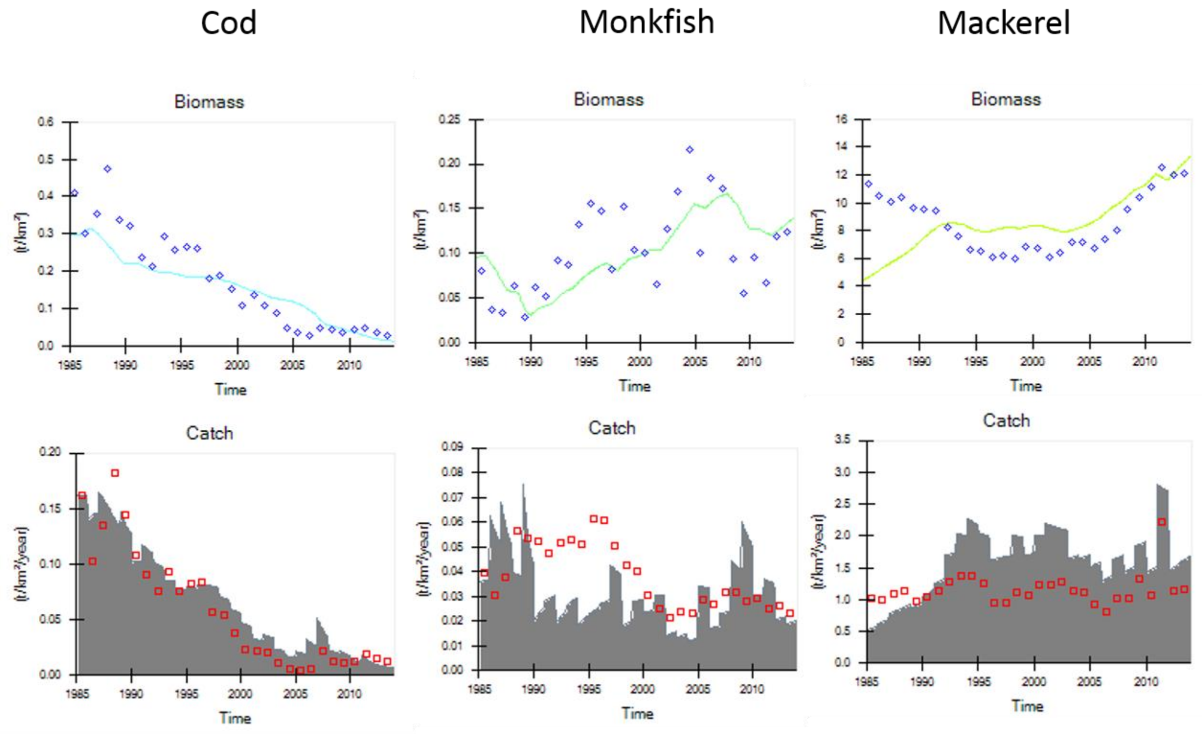

Figure 1: Examples of model estimates (solid lines/histograms) of biomass and catches compared to historical values (dots) for cod, monkfish and mackerel for the West of Scotland EwE model.

Gadget is a statistical modelling environment which can be used to describe the population dynamics of species within a given ecosystem. Models can be implemented on a single species basis (i.e. similar to typical stock assessment models; Fig. 2), or in a multispecies framework where several species interact through predator-prey relationships. Multiple, interacting fishing fleets may also be specified in Gadget models, with each fleet removing predetermined target species. In addition, discards and misreported catch may be modelled. Gadget has the flexibility to incorporate data from many

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

different sources to a single model, including commercial catch-at-age data, scientific survey indices, and stomach content data. Models are fitted to the data using a formal statistical approach, whereby each species' population is simulated given a set of initial conditions, population parameters (e.g. recruitment, fishing mortality) are estimated by maximum likelihood, and the simulated population is compared to the data through negative log-likelihood functions. The fitting procedure itself is objective, but care must be taken to ensure the structure of the model is ecologically sensible. Once a satisfactory model is attained, Gadget can provide forward projections of various population indices based on the parameters of the modelled ecosystem.

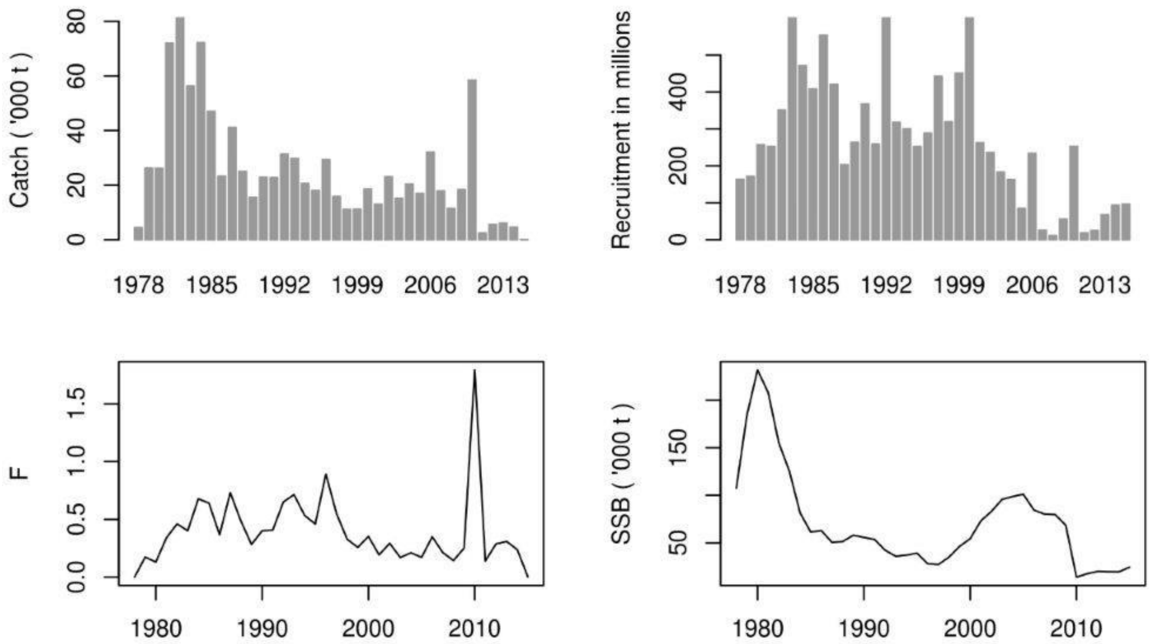

Figure 2: Example of outputs from Gadget

EwE was developed to investigate species interactions in complex foodwebs such as prey/predators feedback loops, competition, etc. As such it encapsulates the whole foodweb from primary producers to top predators (i.e. mammals, seabirds). It also allows for including, and investigating the effects of, abiotic factors such as forcing functions, temperature, salinity, etc. and is a useful tool to do so. As fishing is a major source of mortality it also allows the inclusion of different fishing fleets targeting different groups. However, it was not initially designed to investigate fisheries related issues and is therefore limited in doing so. Major drawbacks include the lack of age/length definition which prevents from investigating the effects of selectivity (e.g. mesh sizes, etc.) and the lack of discards which prevents from investigating choke species issues. On the other hand, because it includes a large number of species EwE allows for a 'true' ecosystem approach: all trophic levels are considered and ecosystem indicators such as biodiversity and mean trophic levels can easily be computed (for full list of indicators see D 4.3). In addition, with all exploited species and corresponding fishing fleets include, it allows for economic outputs such as revenues, etc.

Gadget models simulate populations in terms of species' age/length structure, thus providing population indices similar to 'classic' single-species assessment approaches. The ability to include multiple, interacting fishing fleets in the model, combined with its age/length structured nature, mean that Gadget may be used to investigate issues which are of interest to fisheries management such as fleet selectivity. Gadget also has facility to model the effects on certain biological processes such as growth. Implementing multispecies models in Gadget can become quite complex, and so it is perhaps better suited to modelling more simple (or targeted) ecological interactions (e.g. a mixed species demersal fishery).

Conclusions

EwE allows for a 'true' ecosystem approach by including the whole foodweb and the corresponding ecosystem indicators. By including a large number of species it accounts for complex prey/predator interaction and for knock-on effects of fishing e.g. recovery of preys when predator is fished and vice-versa. While the realism with which the fishing impact is modelled due to lack of age/length classes can be limiting, it provides indications of what is likely to happen in the foodweb when a specific exploitation regime is applied. EwE is unlikely to be employed for short term tactical and technical management decisions where more realistic models are preferred, but it provides insight on the long term directional trends at the ecosystem level which could still be use to inform managers in an EAFM context. Within MareFrame, the lack of discards in the EwE model proved limiting as stakeholders happen to be greatly concern by the short-term implications of the landings obligation.

In contrast, Gadget may be better suited to modelling targeted interactions within an ecosystem which are considered important in terms of their effects on commercial species. The age/length structure of the model, and the objective nature of the fitting procedure mean that it is well suited for the provision of tactical management advice for a given fish stock(s). For the West of Scotland case study, single species Gadget models have been developed for a group of commercially important gadoid species (cod, haddock, whiting), with the catch of each species attributed to landings, discards and misreported catch. Including these processes means that the models will offer insights into current management issues including the landings obligation. The final model will include interactions between the above species in a multispecies context.

The West of Scotland case study in a position where the suitability and performance of two completely different modelling tools i.e. end-to-end foodweb model vs. age-based multispecies model can be assessed when modelling the same ecosystem. Hopefully the MareFrame project can draw useful insight form this comparison, such as the pro and cons of each tool when investigating specific management strategies.

Iberian case study

Stakeholder interactions were concentrated in Gulf of Cadiz ecosystem, giving priority to the anchovy stock. They asked an advice based on adaptive management that integrates the environmental forcing on population dynamics and the socio-economic aspects. In order to achieve this goal we have chosen two models, a minimum realistic model (MRM) described in Rincon et al 2016 and a Gadget model. Table X summarises the application stage of each model within the case study. The MRM is fully implemented in R including ecological and socio-economic indicators for different management scenarios. Using the advantages of a growing platform as R, it was easy to translate the model outputs into a user friendly interface using *shiny* R package (Chang et al. 2016) that was linked to the MareFrame DST website and presented to stakeholders.

The Gadget model has been parameterized and a forecast implemented, but a Gadget model including environmental effects before recruitment is at a preliminary stage.

These two models are not comparable because the first one acts as an operating model to provide time series of simulated abundance, while the other estimates population dynamics parameters based on fisheries data. To use them in a comparison framework, we chose other simple models to contrast with Gadget output. Following a MSE approach, we simulated biomass and catches time series with the operating model to have a reference of how well is the TAC advice provided by Gadget and other models described in the 'MareFrame models at comparison' section.

Table 4 Overview of application stage of each model in the case study

Application stage	MRM	GADGET
Model design completed	✔	✔
Input files for environmental data prepared	✔	✔
Model is fully parametrized	✔	✔
Model creates reasonable/expected results	✔	✔
Output linked to ecological and economic indicators	✔	✗
All management scenarios modelled in satisfactory way	✔	✗
Outputs prepared for input into DST	✔	✗
Outputs presented to stakeholders	✔	✗

All the models estimate Fmsy and the current biomass to calculate the TAC value as the common metric. Calculation of Fmsy changed across methods while biomass was approximated sampling 1000 possible values from a normal distribution with mean equal to estimated current biomass and coefficient of variation equal to that of catch time-series. In particular, for Gadget, the Fmsy was calculated using a stochastic forward simulation (through Gadget.forward function from R package RGadget (Elvarsson, 2015)) of 100 years under different harvest rates. The mean of the normal distribution was calculated as the mean of the estimated biomass for the last 5 years.

Comparison using real data

Two Gadget models were tested, one using data available from 1988 to 2015 and the other from 2001 to 2015. This difference was made because Gadget has a different performance in both periods because of the length-distribution fit. The absence of regulation together with extreme environmental conditions in the first years (1988-2000) produce length-distribution patterns that Gadget does not fit well. This lack of fit in the first Gadget model results in an TAC close that estimated by data limited methods (Figure 3). This suggests that Gadget becomes close to a precautinary approach (such as data limited methods do) for this period. For the second period (2001-2015), the length distribution fit achived by Gadget improves. The resulting TAC in this second peridos becomes less conservative and, cosnquently, higher than data limited methods (Figure 4).

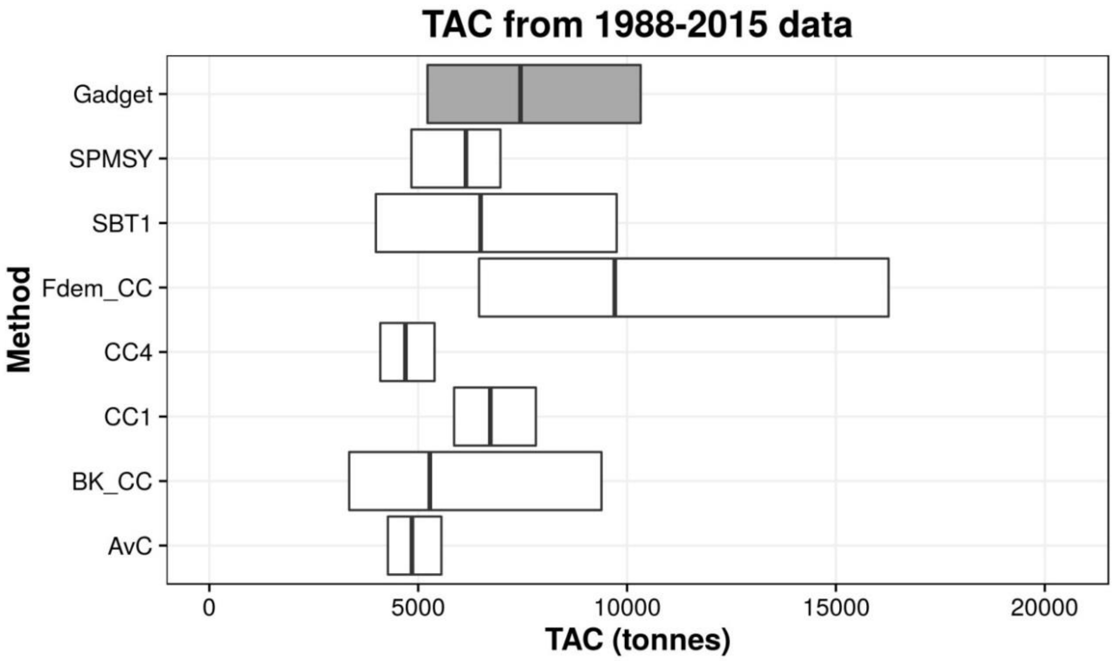

Figure 3: Comparison of estimated TAC value between a Gadget model and other models using information from 1988 to 2015. The upper and lower "hinges" correspond to the first and third quartiles (the 25th and 75th percentiles).

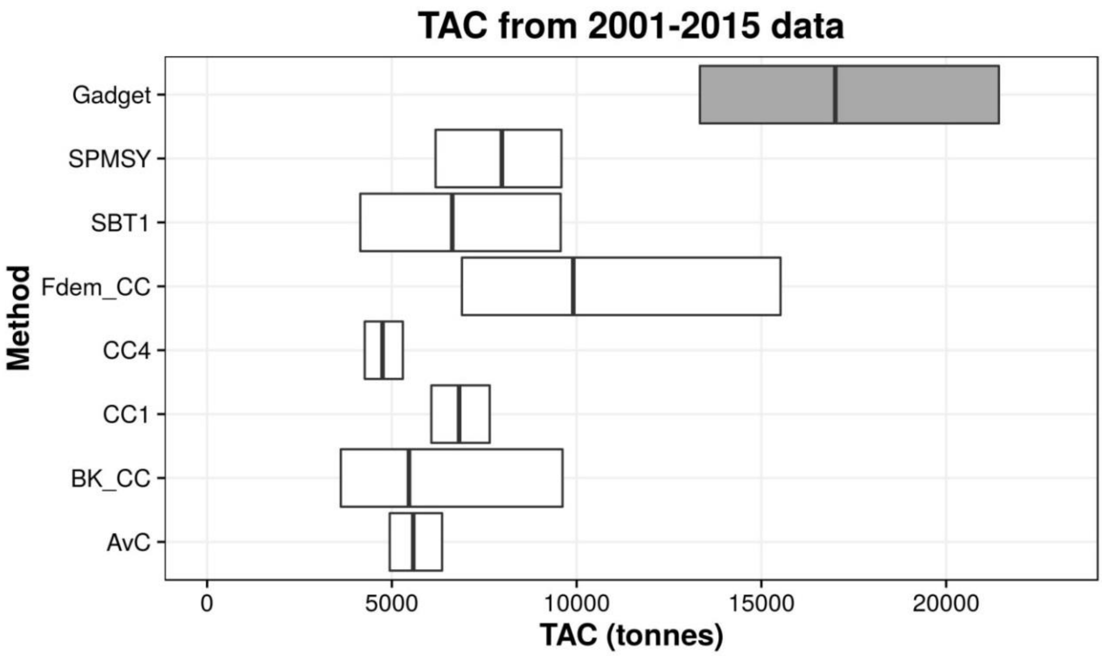

Figure 4: Comparison of estimated TAC value between a Gadget model and other models using information from 2001 to 2015. The upper and lower "hinges" correspond to the first and third quartiles (the 25th and 75th percentiles).

Management Strategy Evaluation comparison framework

The operating model proposed in Rincon *et al.* 2016 simulates the dynamics of this stock for a 30 year period as forced by sea surface temperature (SST), easterly winds and discharges from the Guadalquivir River. This model calculates biomass time-series for 10 simulated populations. In these simulations, the number of spawning events (related to SST), windy days and discharges were randomly sampled from uniform and lognormal distributions. These distributions were chosen based on ranges from historical records of wind and discharges. Considering the seasonal pattern of discharges and the period when the juveniles occupy the estuary, mean, and standard deviation of the logarithm were calculated using only the discharges from March to October of each year. Catches were extracted from these 10 populations using Baranov equation with a target constant fishing mortality. The target monthly fishing mortality (0.04) was chosen to lie below Fmsy. This value was identified by performing simulations with different fishing mortalities to find the one that gave the highest average yield in stochastic simulations over the 30-year period.

For each of the 10 simulated population dynamics, the biomass and catch time-series are defined as the reference, therefore representing the "truth". For comparison purposes, some variability was added to the reference biomass in the last year (current reference-biomass) by sampling 1000 values from a normal distribution. The mean of this distribution was defined as the mean of the last 5 years of reference biomasses while the CV was that of the whole reference-catch time-series. Reference TAC was defined as the product of these values and the target annual fishing mortality (0.04*12).

Reference catches and an ad hoc created "perfect" biomass survey, with the same values as the reference biomass, were included as data to Gadget and the other models. TAC values were calculated as described above except for Fmsy in Gadget where the stochastic forward simulation results in an increasing function of yield versus harvest rate. In this case Fmsy was defined as the F that reduces the initial simulated biomass in a 30%.

The comparison of Gadget and the other models against the reference can be observed in Figures 5 and 6.In most of the simulations the implementation of data limited methods results in a more precautionary TAC than the calculated with Gadget. Gadget TAC is closer to the reference in most of the scenarios but the median is higher in all the cases. This difference is due to the Fmsy calculated by Gadget. This is higher than the target annual fishing mortality (0.04*12) used in the reference population for all the simulations. While Gadget calculates a very good estimate of the biomass time-series, the F is very low and, consequently, the effort that maximize the yield is very high. Even if we change the definition of Fmsy as the F that reduces the initial simulated biomass in a 30%, it is still very high.

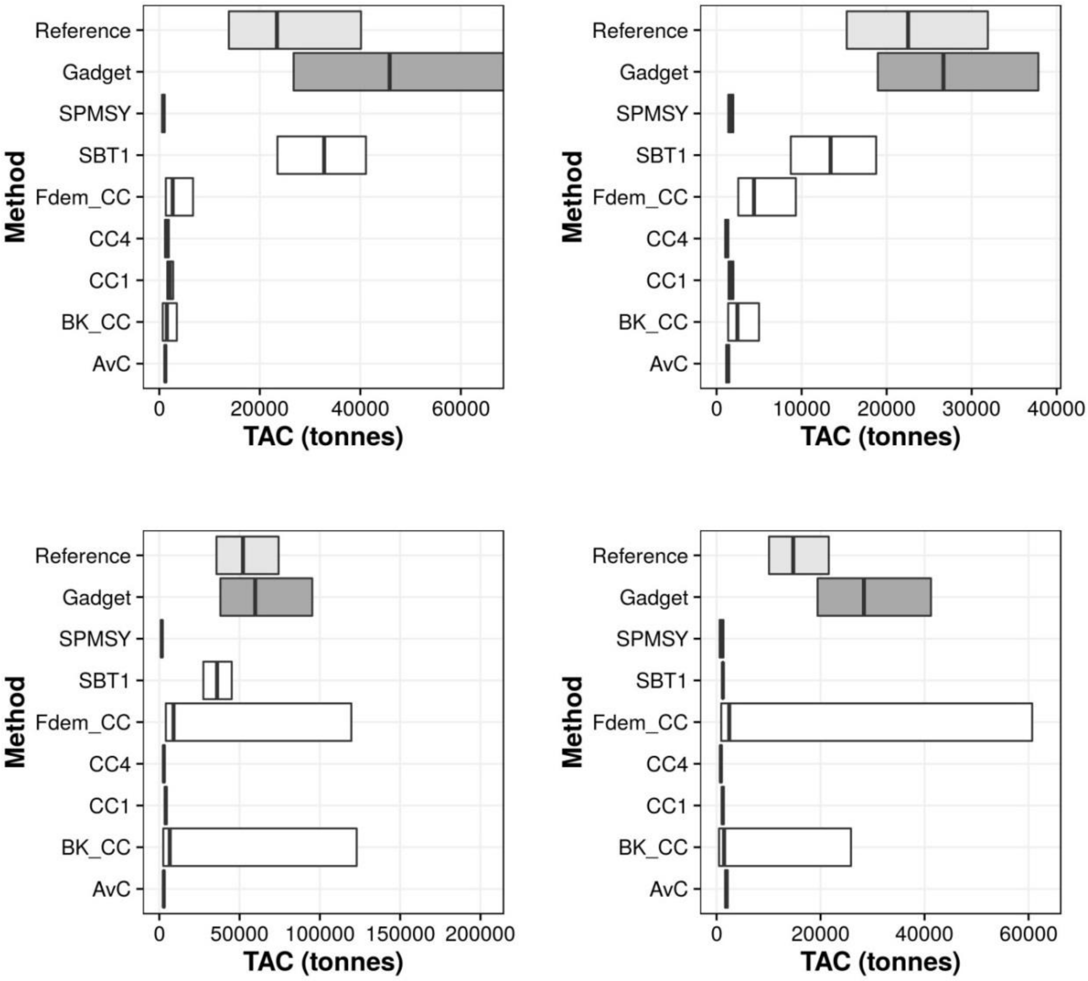

Figure 6: Comparison of TAC values derived from Gadget and other methods with a reference simulated population. Each plot represents a simulation and upper and lower "hinges" correspond to the first and third quartiles (the 25th and 75th percentiles)

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

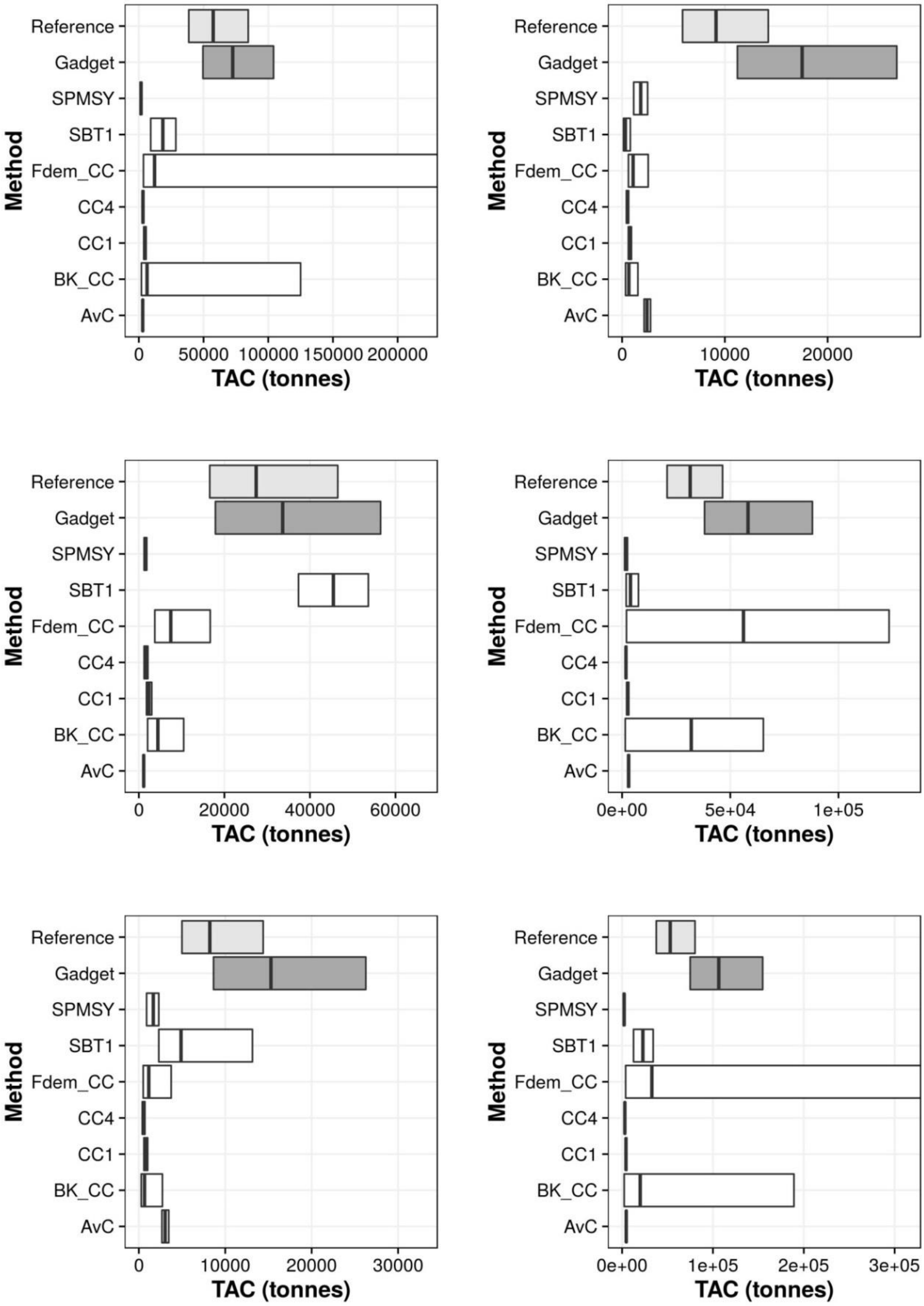

Figure 7: Comparison of TAC values derived from Gadget and other methods with a reference simulated population. Each plot represents a simulation and upper and lower "hinges" correspond to the first and third quartiles (the 25th and 75th percentiles).

Catches in the reference time-series are driven by environmental variability plus a constant fishing rate, but Gadget did not include the enviromental information in this exercise. Therefore, in order to account for this variability, it estimates a fishing mortality that is not constant and with a very smal value. This is consistent with the conclusion reached by Rincon et al. (2016) when stating that, under significant sources of environmentally driven variability, a constant fishing mortality regime will either overexploit a given cohort or under-exploit it. The very low value that Gadget estimates for F reflects that Gadget can get the whole picture of simulated underexploited populations.
This MSE exercise shows that Gadget as an integrated model which is able to use multiple sources of information at the same time, calculates better estimates than the other methods in all simulated populations. These are preliminary results from a manuscript that includes a broader comparison requiring more computational time and that will be available soon.

Conclusions
The operating model accounts for the environmental effect on early anchovy stages. Further it has been expanded with a socio-economic module (Deliverable 4.5 and Ruiz et al, *submitted*), including socio-economic and ecological indicators such as gross value added, labour cost, full time employment and collapse probability. Since these indicators are calculated for the fishery and also by vessel, the information becomes attractive to stakeholders and consequently this model has been effectively used for the DST implementation. It also allows to answer what-if questions like, what if we insure the fishery? What if we use an environmentally-based harvest control rule? This model can replicate stock trends, and that is the reason why we use it also for simulations of the "truth" to compare with other models. It is a versatile tool to see the global landscape but the lack of an estimation method makes it inadequate for short-term assessments.
On the other hand, Gadget performance compared with the other methods was better in all the simulations; it replicates abundance indexes and catches time series with great accuracy. As we said before, the Gadget model used in this exercise does not include the environmental information but we are working to include the environmental variability as data for Gadget through the spawning-recruitment relationship (Deliverable 5.3). Another option would be to use the relationship between the estimated recruitment by Gadget and the environment in the forecast. Both approaches would be tested again in a MSE framework to see how this input affects the fishing mortality and Fmsy estimations. This exercise confirms that Gadget is the best choice in order to give an advice for management purposes due to its the statistical core and the flexibility to include ecosystem processes.

Mediterranean case study

The Strait of Sicily (SoS) case study (CS) focuses on the development of a reliable tool for the implementation of the ecosystem approach to fisheries management (EAFM) in a key fishing area in the Mediterranean Sea. The objectives of the CS have been progressively refined through the application of a co-creation approach with key stakeholders (i.e. fishers and fishers representatives, managers of local and national administrations, conservation NGOs, FAO and GFCM officers) and taking into account the objectives of the GFCM international management plan for bottom trawl fisheries exploiting deep water rose shrimp (DPS: *Parapenaeus longirostris*) and hake (HKE: *Merluccius*

merluccius) in GSAs 12-16 (i.e. Strait of Sicily, GFCM, 2016[4]). These fisheries are the most important in the SoS region both from a socio-economic point of view and considering their impacts on the ecosystem. Trawlers from different nations (i.e. Italy, Tunisia, Malta, Libya, Egypt) and under different management regimes exploit shared stocks in national and international waters thus making challenging the implementation of agreed management rules.

In particular the GFCM plan includes the establishment of three FRAs (Fisheries Restricted Areas) and the closure of the Gulf of Gabes (GSA 13) for three months in summer. It also establishes the objective to achieve F_{msy} for HKE and DPS by 2020.

The overall goal of the CS has been adapted to provide a tool for the application of EAFM in the SoS which can support the achievement of long term sustainability by finding a balance between ecological and human well-being through good governance. In turn the CS also might also substantially contribute to the development of the GFCM management plan through the inclusion of a more holistic approach. Atlantis (Fulton et al., 2004) and Gadget ecosystem models have been parameterized to provide advice on the effects of different management scenarios in respect of the following four main management objectives identified during the case study meetings: i) rebuilding overexploited stocks; ii) long-term continuity of the fishing activities; iii) same rules for all; iv) good environmental status. Atlantis is a complex ecosystem model able to represent the high complexity of the Mediterranean ecosystem and forecast the impact of management measure as well as climate forcing on key ecosystem processes, functional groups, populations and fisheries (Fulton et.al, 2004). Gadget is a parametric forward-simulation model of an ecosystem, typically consisting of various fish populations, fleets and their interactions. Plagányi (2007) has classified Gadget as a “Minimum Realistic Model (MRM)” to describe the concept of restricting a model to those species most likely to have important interactions with the species of interest. On the other hand, Atlantis allows to explore the effects on target stocks as well as the other functional groups included in the model providing a more holistic view on the impacts of the simulated management measures applied. In addition, Atlantis is spatially structured thus allowing to explore the effects of spatially-based management measures. The two models have been customized to provide management advice in the region on the basis of the advice provided by the stakeholders during 3 CS meetings organized in close collaboration with WP1 and WP 6. The application stage of the two models is summarized in table 6.

[4] GFCM, 2016. REC.CM-GFCM/40/2016/4 establishing a multiannual management plan for the fisheries exploiting European hake and deep-water rose shrimp in the Strait of Sicily (GSA 12 to 16)

Table 6: Overview of application stage of each model in the case study

Application stage	ATLANTIS	GADGET
Model design completed	✔	✔
Input files for environmental data prepared	✔	✘
Model is fully parametrized	✔	✔
Model creates reasonable/expected results	✔	✔
Output linked to ecological and economic indicators	✔	✘
All management scenarios modelled in satisfactory way	✘	✘
Outputs prepared for input into DST	✔	✘
Outputs presented to stakeholders	✘	✘

The two models estimate fishing mortality and biomass of hake and deep-water rose shrimp, and are able to simulate catches under different management scenarios.

Atlantis is composed by a set of submodels. It features a deterministic biophysical submodel, which is spatially resolved in three dimensions using a map made up of polygons and vertical layers. It follows tracks the nutrient flow through the main biological groups found in the marine ecosystem of interest. The primary ecological processes considered in the model are consumption, production, waste production and cycling, migration, predation, recruitment, habitat dependency, and mortality. Lower trophic levels (invertebrates) are modelled as biomass pools, while the vertebrates are represented using an age- and stock-structured formulation. Physical forcing fields (currents, temperature and salinity) are included using results of an external hydrodynamic model. The exploitation submodel allows for multiple fleets and change in effort allocation.

In this regard Atlantis allows exploration of the multiple effects of human (i.e. fishing pattern) and environmental (i.e. climate forcing) drivers on the dynamics of target stocks and functional groups. Gadget model provide an accurate description on the interactions between hake, deep-water rose shrimp and horse mackerel. It includes Italian and Tunisian trawl and small-scale fleets and account for hake cannibalism. The first two stocks are the target of the GFCM management plan whereas horse mackerel is the key foraging pelagic species in the region.

Both Gadget and Atlantis can provide F_{MSY} and F-at-age estimates, stock biomass and forecast catch (see D. 5.3.). In both models food consumption is explicitely included, limited in Gadget to the consumption of DPS and HOM by hake length groups. Atlantis might incorporate the dependency of key stocks and functional groups on environmental drivers which are not yet represented in our Gadget model.

In the current formulation Gadget is more suitable to account for fisheries catchability than Atlantis, but on the other hand Atlantis can dynamically model the spatial distribution of fishing effort/catches and fishing mortality providing a robust tool to investigate the impact of spatial management measures (e.g. area closures/MPA).

Economic and social indicators are modelled through assumed relationships between fishing days and fishing mortality. Costs, revenues and profits are calculated on the basis of the estimated landings/cpue under different management scenarios.

Table 7: Summary of relevant ecosystem components for magament and whether they are modelled within the modelling approaches.

Management considerations	ATLANTIS	GADGET
spawning biomass	✔	✔
size structure	✖	✔
age structure	✔	✔
catches/cpues	✔	✔
biomass target stocks	✔	✔
F/Fmsy	✔	✔
by-catch of commercial species (i.e. landing obbligation and minimization of unwanted catches)	✔	✔
non-commercial key food web components as in MSFD (fast-turnover groups, charismatic species)	✔	✖
ecosystem indicators	✔	✖
effect of environmental forcing on target stocks and key functional groups	✔	✖
social and economic effects of the simulated measures	✔	✔

Black Sea case study

The Black Sea ecosystem is seriously affected by dynamic changes directly related to fishing, climate change and pollution. Fishery is the sector most affected by the changes of the Black Sea ecosystem. In the same time, fishing activities contribute themselves to the worsening of the ecological situation and for the depletion of the fish stocks.The objective of the Black Sea case study is the restoration of turbot fisheries to more productive levels, considering both the effect of fisheries and the ecosystem change that has occurred in the last 30 years. The ecosystem models employed in this case of study are Gadget and EwE, with the aim of increasing the knowledge about the Black Sea ecosystem functioning and thereby serve to advise on the rebuilding of the turbot stock. These models will allow providing input to the development of a management plan.

Black Sea Case Study - status of work

☒ Model design completed

☒ Input file prepared

☒ Model is fully parametrized

☒ Model creates reasonable/expected results

☒ Side models / accessory data completed

☒ Model work fully completed

Ecopath with Ecosim (EwE)

For develop our scenarios, we consider three kind of measures:

Business As Usual = 100 % IUU

Soft Measures = 50 % IUU

Hard Measures = NO IUU

And three kind of Harvest Control Rule:

Fishing Mortality(F)

Total Allowable Catch(TAC)

Maximum Sustainable Yield(MSY)

Dataset used is from West Black Sea, i.e. from Romania, Bulgaria, Ukraina. The desires output: SSB, catches, landing.

For variable F: Run Ecosim → ecosim group plots → check turbot at age 2,3,4,5,6,7, where F exist → save data to csv → average of F/year

For variable TAC: we consider TAC = an imposed catch. Next, we make a new CSV file with TAC instead of catches → load CSV → Run Ecosim → Ecosim group plots

For MSY: $MSY = 0.37 \times M \times B_{med}$

	INPUT				OUTPUT		
	Year	BAU	F	Year	SSB	Catch	Landing
Scenario 1	2007	0.093241	0.208766	2014	0.016491	0.001103	0.001103
	2008	0.099309	0.208766	2015	0.017076	0.001031	0.001031
	2009	0.101498	0.22289	2016	0.031744	0.002208	0.002208
	2010	0.075984	0.235241	2017	0.050939	0.003033	0.003033
	2011	0.085121	0.145753	2018	0.066313	0.003718	0.003718
	2012	0.088203	0.191784	2019	0.07561	0.003929	0.003929
	2013	0.07689	0.154388	2020	0.075559	0.003905	0.003905
	Year	BAU	TAC	Year	SSB	Catch	Landing
Scenario 2	2007	0.093241	0.002915	2014	0.048091	0.002915	0.002507
	2008	0.099309	0.002915	2015	0.048714	0.002915	0.002507
	2009	0.101498	0.002507	2016	0.049104	0.002507	0.002507
	2010	0.075984	0.002507	2017	0.049347	0.002507	0.002507
	2011	0.085121	0.002507	2018	0.049491	0.002507	0.002507
	2012	0.088203	0.002507	2019	0.049547	0.002507	0.002507
	2013	0.07689	0.002507	2020	0.049545	0.002507	0.002507
	Year	BAU	MSY	Year	SSB	Catch	Landing
Scenario 3	2007	0.093241	0.071852	2014	0.957669	1.316489	1.316489
	2008	0.099309	0.068232	2015	0.955221	1.281745	1.281745
	2009	0.101498	0.066621	2016	0.953242	1.266618	1.266618
	2010	0.075984	0.065506	2017	0.951282	1.246292	1.246292
	2011	0.085121	0.064865	2018	0.94905	1.236737	1.236737
	2012	0.088203	0.064177	2019	0.946102	1.214033	1.214033
	2013	0.07689	0.063144	2020	0.937969	1.210515	1.210515

Figure.1: Business As Usual

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613571

	INPUT					OUTPUT		
	Year	SM	F		Year	SSB	Catch	Landing
Scenario 4	2007	0.046621	0.034794		2014	0.053982	0.003108	0.003108
	2008	0.049655	0.034794		2015	0.054383	0.003108	0.003108
	2009	0.050749	0.036952		2016	0.054463	0.003469	0.003469
	2010	0.037992	0.039035		2017	0.054393	0.003886	0.003886
	2011	0.04256	0.025535		2018	0.054289	0.002626	0.002626
	2012	0.044101	0.031325		2019	0.054167	0.003465	0.003465
	2013	0.038445	0.026251		2020	0.054037	0.002911	0.002911
	Year	SM	TAC		Year	SSB	Catch	Landing
Scenario 5	2007	0.046621	0.002915		2014	0.053982	0.002915	0.002507
	2008	0.049655	0.002915		2015	0.054383	0.002915	0.002507
	2009	0.050749	0.002507		2016	0.054463	0.002507	0.002507
	2010	0.037992	0.002507		2017	0.054393	0.002507	0.002507
	2011	0.04256	0.002507		2018	0.054289	0.002507	0.002507
	2012	0.044101	0.002507		2019	0.054167	0.002507	0.002507
	2013	0.038445	0.002507		2020	0.054037	0.002507	0.002507
	Year	SM	MSY		Year	SSB	Catch	Landing
Scenario 6	2007	0.046621	0.055926		2014	0.745408	0.888084	0.888084
	2008	0.049655	0.052985		2015	0.741766	0.908591	0.908591
	2009	0.050749	0.051581		2016	0.738039	0.92847	0.92847
	2010	0.037992	0.05056		2017	0.734224	0.947719	0.947719
	2011	0.04256	0.049915		2018	0.730322	0.966334	0.966334
	2012	0.044101	0.049269		2019	0.726331	0.984313	0.984313
	2013	0.038445	0.048621		2020	0.722249	1.001652	1.001652

Figure.2: Soft Measures

	INPUT				OUTPUT			
	Year	HM	F		Year	SSB	Catch	Landing
Scenario 7	2007	0	0.005799		2014	0.075451	0.00338	0.00338
	2008	0	0.005799		2015	0.07498	0.003366	0.003366
	2009	0	0.006126		2016	0.074496	0.003348	0.003348
	2010	0	0.006477		2017	0.074077	0.003336	0.003336
	2011	0	0.004474		2018	0.073835	0.003327	0.003327
	2012	0	0.005116		2019	0.073544	0.003315	0.003315
	2013	0	0.004463		2020	0.073265	0.003311	0.003311
	Year	HM	TAC		Year	SSB	Catch	Landing
Scenario 8	2007	0	0.002915		2014	0.058875	0.002915	0.002507
	2008	0	0.002915		2015	0.0586	0.002915	0.002507
	2009	0	0.002507		2016	0.058277	0.002507	0.002507
	2010	0	0.002507		2017	0.058015	0.002507	0.002507
	2011	0	0.002507		2018	0.057801	0.002507	0.002507
	2012	0	0.002507		2019	0.05759	0.002507	0.002507
	2013	0	0.002507		2020	0.057382	0.002507	0.002507
	Year	HM	MSY		Year	SSB	Catch	Landing
Scenario 9	2007	0	0.053873		2014	0.718047	0.827671	0.827671
	2008	0	0.050996		2015	0.713931	0.820807	0.820807
	2009	0	0.049601		2016	0.709713	0.81381	0.81381
	2010	0	0.048574		2017	0.705391	0.806682	0.806682
	2011	0	0.047909		2018	0.700966	0.799422	0.799422
	2012	0	0.047241		2019	0.696435	0.792032	0.792032
	2013	0	0.046571		2020	0.691797	0.784512	0.784512

Figure 3: Hard Measures

Gadget

For Gadget we use same scenarios as for EwE.

	INPUT				OUTPUT		
Scenario 1	Year	BAU	F		SSB	Catch	Landing
	2007	0.093241	0.208766		0.016577	0.00162	0.00162
	2008	0.099309	0.208766		0.017643	0.007474	0.007474
	2009	0.101498	0.22289		0.03524	0.003446	0.003446
	2010	0.075984	0.235241		0.01626	0.001589	0.001589
	2011	0.085121	0.145753		0.07506	0.007337	0.007337
	2012	0.088203	0.191784		0.03469	0.00339	0.00339
	2013	0.07689	0.154388		0.01608	0.00157	0.00157
Scenario 2	Year	BAU	TAC		SSB	Catch	Landing
	2007	0.093241	0.002915		0.048342	0.004724	0.004724
	2008	0.099309	0.002915		0.050332	0.021322	0.021322
	2009	0.101498	0.002507		0.054513	0.005331	0.005331
	2010	0.075984	0.002507		0.015752	0.001539	0.001539
	2011	0.085121	0.002507		0.056019	0.005476	0.005476
	2012	0.088203	0.002507		0.022732	0.002221	0.002221
	2013	0.07689	0.002507		0.010544	0.00103	0.00103
Scenario 3	Year	BAU	MSY		SSB	Catch	Landing
	2007	0.093241	0.071852				
	2008	0.099309	0.068232				
	2009	0.101498	0.066621				
	2010	0.075984	0.065506				
	2011	0.085121	0.064865				
	2012	0.088203	0.064177				
	2013	0.07689	0.063144				

Figure 4: Business As Usual

	INPUT				OUTPUT		
Scenario 4	Year	SM	F		SSB	Catch	Landing
	2007	0.046621	0.034794		0.054263	0.004567	0.004567
	2008	0.049655	0.034794		0.05619	0.022533	0.022533
	2009	0.050749	0.036952		0.060462	0.005415	0.005415
	2010	0.037992	0.039035		0.017363	0.002036	0.002036
	2011	0.04256	0.025535		0.06145	0.005181	0.005181
	2012	0.044101	0.031325		0.024852	0.00299	0.00299
	2013	0.038445	0.026251		0.0115	0.00117	0.00117
Scenario 5	Year	SM	TAC		SSB	Catch	Landing
	2007	0.046621	0.002915		0.054263	0.002507	0.002507
	2008	0.049655	0.002915		0.05619	0.002507	0.002507
	2009	0.050749	0.002507		0.060462	0.002507	0.002507
	2010	0.037992	0.002507		0.017363	0.002507	0.002507
	2011	0.04256	0.002507		0.06145	0.002507	0.002507
	2012	0.044101	0.002507		0.024852	0.002507	0.002507
	2013	0.038445	0.002507		0.0115	0.002507	0.002507
Scenario 6	Year	SM	MSY		SSB	Catch	Landing
	2007	0.046621	0.055926				
	2008	0.049655	0.052985				
	2009	0.050749	0.051581				
	2010	0.037992	0.05056				
	2011	0.04256	0.049915				
	2012	0.044101	0.049269				
	2013	0.038445	0.048621				

Figure 5: Soft Measures

	INPUT				OUTPUT		
Scenario 7	**Year**	**HM**	**F**		**SSB**	**Catch**	**Landing**
	2007	0	0.005799		0.075844	0.004966	0.004966
	2008	0	0.005799		0.077471	0.0244	0.0244
	2009	0	0.006126		0.082701	0.005225	0.005225
	2010	0	0.006477		0.023646	0.001748	0.001748
	2011	0	0.004474		0.083575	0.006565	0.006565
	2012	0	0.005116		0.033742	0.002861	0.002861
	2013	0	0.004463		0.015592	0.001331	0.001331
Scenario 8	**Year**	**HM**	**TAC**		**SSB**	**Catch**	**Landing**
	2007	0	0.002915		0.059182	0.002507	0.002507
	2008	0	0.002915		0.060548	0.002507	0.002507
	2009	0	0.002507		0.064696	0.002507	0.002507
	2010	0	0.002507		0.018519	0.002507	0.002507
	2011	0	0.002507		0.065425	0.002507	0.002507
	2012	0	0.002507		0.026422	0.002507	0.002507
	2013	0	0.002507		0.012212	0.002507	0.002507
Scenario 9	**Year**	**HM**	**MSY**		**SSB**	**Catch**	**Landing**
	2007	0	0.053873				
	2008	0	0.050996				
	2009	0	0.049601				
	2010	0	0.048574				
	2011	0	0.047909				
	2012	0	0.047241				
	2013	0	0.046571				

Figure 6: Hard Measures

Pros and cons

Management questions	EwE	GADGET
Biomass trend(total and SSB)	Y	Y
F as against F_{MSY}	Y	N
impact of turbot stock state to all species state from food web	Y	Y
influence of SSB on recruitment	Y	Y
influence degree of environmental factors on recruitment and production	Y	Y
setting the fishing tool type with maxim impact on state of stock	Y	N
influence degree of management measures, like:		
1)increase time of prohibition	N	N
2)increase the forbiden area for fishing	Y	N

In conclusion, in classic turbot management in the Black Sea, only tactical decision issues have been considered (the business strategy) so far. This is the first time when by using Gadget and EwE, the decision-making process also considers environmental,stock structure and socio-economic aspects, which should support taking the best strategic decisions for managing this particularly valuable and vulnerable species.

Conclusions

- There are interconnections amongst the components of any ecosystem. Therefore taking account of biological interactions, such as predator-prey interactions, in long-term management decisions may substantially change perspectives and the way marine resources are used and managed. This applies at the level of both individual fish populations and of entire marine ecosystems. It has resulted in an increase in the demand for the development and application of ecosystem models where these are able to characterise the trade-offs amongst different management objectives. An ability to take these trade-offs into account is a central aspect of the ecosystem approach to fisheries management (EAFM).
- For all of the case studies considered, the ecosystem models examined show considerable variability in their output. This result was not unexpected, given the high structural uncertainty inherent in ecosystem models (Fulton et al. 2003, Link et al. 2012). Some of these differences arise from the wide range in scope (from tactical to strategic) covered by the different models examined, as well as from the different extents to which they focused on securing good fits to the data available. In general, comparative approaches are recommended as the way forward, both to quantify structural uncertainty and to find results which are robust to model formulation (Forrest et al. 2015).
- All the MareFrame case studies adopted the co-creation approach. This led to confirmation of the high potential which ecosystem models possess to highlight the trade-offs to which fisheries management needs to give attention. However, in several cases the approach also served to emphasise the limits of the current models and the difficulties in implementing them to address some of the specific '*co-created management objectives*'. Limitations were evident in their ability to simulate certain scenarios and to address forecasting requests. For example, difficulties in implementing multi-area ecosystem models restricts their ability to address the effects of MPAs or FRAs. Furthermore, predictions of the effects of trophic cascades arising from rebuilding populations of top predators (e.g. seals, cod, etc.) are highly sensitive to the complex predator-prey interactions that take place at high trophic levels and are not fully understood. The co-creation approach may require ecosystem models that are sufficiently flexible to address new objectives or test alternative management strategies from time to time, when such aspects may arise during consultations with stakeholders. Often the ecosystem models considered were found to lack such flexibility. However, experience from different case studies has shown that the use of multiple models of increasing complexity (e.g. from Gadget, to EwE, to Atlantis) can partially address this issue.
- A general feature of the ecosystem models considered is that increased model complexity comes at the expense of precision and ability to fit available data. The inclusion of more species, trophic layers and processes often requires more assumptions, readily finds itself compromised by paucity of data, and can lead to difficulties in achieving statistically appropriate fits to data. Nevertheless, the management questions posed, and the development of the associated decision support tools, for the different case studies were found to require models which addressed certain aspects of this increased ecosystem complexity. These aspects go far beyond what traditionally needs to be considered for single species stock assessments.
- The various ecosystem models considered in the case studies have been described in an earlier section. Models such as the Charmingly Simple (CSM) or Multispecies Schaefer are relatively simple

compared to EwE, Atlantis and Gadget. However, they have the advantage of being user friendly and are better suited for interactions with stakeholders. The Green Model, which summarizes more complex models, e.g. SMS (the ICES Stochastic Multispecies Model) or EwE, uses a multispecies Schaefer model to approximate the behaviour of the more complex models that it seeks to emulate. This means that computer runs complete quickly and this, together with the Green Model's added ability to bolt on social, economic and GES modules, makes it very well suited to use in interactions with stakeholders. Some simple models may also be appropriate in data-limited situations (e.g. simple models in south western waters). All these models are potentially valuable additions to the MareFrame tool box since they can be useful for different purposes.

Both EwE and Atlantis are 'true' ecosystem models and do have the strength of including a large number of species. They can therefore be employed as strategic tools to compare scenarios, explore trade-offs, and ultimately inform management decisions (i.e. by predicting what is likely to happen) both for the species included the models and for the overall health of the ecosystem. In addition, Atlantis can function as an operating model to generate simulated data in data-poor situations. Gadget, however, is more suitable for tactical purposes. It is an assessment model which is currently employed to provide some assessments for ICES, and as such can be used to estimate future levels of biomass under a given management scenario. The other models mentioned above tend to be more strategic than tactical in their usage, being simpler user-friendly tools which provide stakeholders with a way of understanding and comparing different management scenarios.

- Regardless of the fact that a management question may require the provision of analyses to inform a short term tactical decision, it is the long-term implications of that decision for the ecosystem are of most interest in an EAFM context. In several of the case studies, rather than indications of which model outperformed the others, what emerged was the need for complementarity. Management of fisheries requires explicit recognition of the complexity of individual fish populations in terms of their abundance and demographic structure, but this does impose strong limitations in the context of an EAFM unless this is limited to a handful of the most important targeted species.
- The case studies have suggested several approaches to model comparison and evaluation. The Management Strategy Evaluation (MSE) approach, better known for the evaluation of approaches for management using simulations (Punt et al. 2014), can also be used to determine how well an estimation model performs, as for example in the Iberian case study. The MSE framework first develops an operating model that provides a mathematical representation of the underlying dynamics of the resource and fishery to simulate the situation under consideration and the data which would typically become available for use in assessments; next an estimation model is used to assess the state of the stock to inform the application of a management strategy; and finally a management strategy option which provides the best performance in terms of objectives is chosen. In the Iberian case study, only the first two steps were implemented through considering a fixed strategy (business as usual) to evaluate how well the estimation models were able to reflect the status of the stock and the fishing pressure. The MSE approach provides a powerful and user-friendly tool for model comparison where the differences of an index estimated from the data relative to the some "reference" level for that indicator can be displayed in a single and simple plot.

  A more complex example offered by MareFrame of the application of simulation models is provided by the use of Atlantis as an operating model (D4.4). The simulated data were fed into two other models - here EwE and Gadget - and their performance was tested and compared in circumstances where the true biomass as represented in the simulated ecosystem, i.e. Atlantis, is known.
- Stakeholder involvement requires that the results of the complex models required for EAFM are packaged in a comprehensible way that enables those stakeholders to understand the trade-offs amongst their specific objectives and concerns. The DSTs used in MareFrame seek to meet this requirement using several different approaches including the following.

- The Multi-Criteria Analysis (MCA) and Bayesian Belief Networks (BBN) provide information on the outcomes of user-defined scenarios. This information may arise from several different models. Potentially this allows for the combination of outcomes from complementary models. Ideally, information and conclusions should be synthesised from a suite of models in a way that allows the stakeholders to take structural uncertainty and/or environmental variability into account in their choices.
- The Green Model offers an example of transportable, nimble and user-friendly representation of the trade-offs of a multispecies multi-fleet fisheries system by approximating the results of more complex, slower-running and less accessible multispecies models.

For all DST developments, the multispecies modeling involves complex biological interactions. These underlie the trade-offs in performance for different objectives and are therefore very relevant. However, these interactions are not trivial to understand, nor to combine across models (e.g. because non-linear behaviour often occurs). Hence, while the DSTs developed in MareFrame have succeeded in linking complex ecosystem models to context-defined user needs, the reliability of their outcomes is not always clear and warrants special attention. The primary use for these DSTs may prove to be in the phases of scoping problems and suggesting potential solutions.

References

Andersen, K. P. and Ursin, E. 1977. A multispecies extension of the Beverton and Holt theory of fishing, with account for phosphorus circulation and primary production. Meddelelser fra Danmarks Fiskeri- og Havundersøgelser, 7: 319-435

CEC 2014: Directive, Strategy Framework. "Directive 2008/56/EC of the European Parliament and of the ouncil." *Journal). Council Decision of* (2008).

Collie, J. S., Gislason, H., and Vinther, M. 2003. Using AMOEBAs to display multispecies, multifleet fisheries advice. – ICES Journal of Marine Science, 60: 709–720.

Collie, Jeremy S., et al. "Ecosystem models for fisheries management: finding the sweet spot." *Fish and Fisheries* (2014).

Cook, R. M., Holmes, S. J., Fryer, R. J. 2015. Grey seal predation impairs recovery of an overexploited fish stock. doi: 10.1111/1365-2664.12439

Elvarsson B. P., Taylor L., Trenkel Verena, Kupca V., Stefansson G. 2014. A bootstrap method for estimating bias and variance in statistical fisheries modelling frameworks using highly disparate datasets. African Journal Of Marine Science, 36: 99-110Elvarsson, B., 2015. RGadget: a R-package for development, testing and analysis of Gadget models. Proceedings of ICES 2015 Annual Science Conference.

Essington T.E., Plaganyi E.E. 2014. Pitfalls and guidelines for "recycling" models for ecosystem-based fisheries management: evaluating model suitability for forage fish fisheries. ICES J. Mar. Sci. 71: 118-127.

FAO (2008) Best practices in ecosystem modelling for informing an ecosystem approach to fisheries. FAO Fisheries Technical Guidelines for Responsible Fisheries. No. 4, Suppl. 2, Add. 1, Rome, FAO, 1–78.

Garcia, S.M.; Zerbi, A.; Aliaume, C.; Do Chi, T.; Lasserre, G. 2013. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. FAO Fisheries Technical Paper. No. 443. Rome, FAO. 2003. 71 p.

Forrest et al., 2015. DO marine ecosystem models give consistent policy evaluations? A comparison of Atlantis and Ecosim. Fisheries Research 167: 293-312.

Francis et al. 2007

Fulton et al. 2003. Mortality and predation in ecosystem models: is it ipmortant how these are expressed? Ecol. Model. 169:157-178.

Fulton, E.A. 2010. Approaches to end-to-end ecosystem models. *Journal of Marine Systems* 81:171–183

Fulton and Smith, 2004. Lesson learnt from a comparison of three ecosystem models for Port Philip Bay, Australia. S. Afr. J. Mar. Sci 26 :219-243.

Fulton, E.A., Smith, A.D.M., Punt,A.E., 2005.Which ecological indicators can robustly detect effects of fishing? ICES J. Mar. Sci. 62, 540–551.

Fulton, E.A. Smith A.D.M., Smith D.C. and Johnson P. 2014. An integrated approach Is needed for Ecosystem Based Fisheries Management: Insights from ecosystem-level Management Strategy Evaluation. *PLoS One*.

Gascuel et al., 2014. Fishing impact and environmental status in European seas: a diagnosis for stock assessments and ecosystem indicators. Fish and Fisheries 17: 31-55.

Harwood J, Stokes K (2003) Coping with uncertainty in ecological advice: lessons from fisheries. Trends Ecol Evol 18:617-622

Hilborn, Ray; Stokes, Kevin; Maguire, Jean-Jacques; Smith, Tony; Botsford, Louis W; Mangel, Marc; Orensanz, JosÃ©; Parma, Ana; Rice, Jake; Bell, Johann; ,When can marine reserves improve fisheries management?,Ocean & Coastal Management,47,3,197-205,2004,Elsevier

Hollowed et al. 2011. Experience with quantitative ecosystem assessment tools in the northeast Pacific. Fish and Fisheries 12: 189-208.

Horbowy J. 2005. The dynamics of Baltic fish stocks based of a multispecies stock production model. J. Appl. Ichthyol.. 21:198-204

ICES. 1992. Report of the Multispecies Assessment WorkingGroup. ICES, *C.M.1992/Assess:16.*

ICES. 1994. Report of the Multispecies Assessment WorkingGroup. ICES, . *C.M.1994/Assess:9*

ICES, 2009. Report of the Working Group on Multispecies Assessment Methods (WGSAM)

ICES, 2010. Report of the Working Group on Multispecies Assessment Methods (WGSAM)

ICES. 2014. Interim Report of the Working Group on Multispecies Assessment Methods (WGSAM), 20–24 October 2014, London, UK. ICES CM 2014/SSGSUE:11. 104 pp.

ICES. 2015. Report of the Workshop on guidance for the review of MSFD decision descriptor 3 – commercial fish and shellfish II (WKGMSFDD3-II), 10-12 February 2015, ICES Headquarters, Denmark. ICES CM 2015\ACOM:48. 36 pp.

Kaplan et al., 2013. Impacts of depleting forage species in the California Current. Environ .Conserv. 40: 380-393.

Laevastu, Taivo, Patricia A. Livingston, and Karl Niggol. *Marine mammals in fisheries ecosystem in the eastern Bering Sea and in the northeastern Pacific Ocean*. Northwest and Alaska Fisheries Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 1980.

Levin et al. 2009. Integrated ecosystem-assessments: developing the scientific basis for ecosystem-based management of the ocean. PlosBiology 7: e1000014.

Lewy, P., M. Vinther A stochastic age-length-structured multispecies model applied to North Sea stocks- ICES CM, 2004 http://www.ices.dk/sites/pub/CM%20Doccuments/2004/FF/FF2004.pdf

Link, J. S. 2010. Adding rigor to ecological network models by evaluating a set of pre-balance diagnostics: A plea for PREBAL. Ecological Modelling 221 (2010) 1580–1591

Longo C., Hornborg S., Bartolino V., Tomczak M.T., Ciannelli L., Libralato S., Belgrano A. 2015. Role of trophic models and indicators in current marine fisheries management. Mar. Ecol. Prog. Ser. 538: 257-272.

Mackinson S & Daskalov G. 2007. An ecosystem model of the North Sea for use in research supporting the ecosystem approach to fisheries management: description and parameterisation. (CEFAS, Lowestoft), p 200. www.cefas.co.uk/publications/techrep/tech142.pdf

McElhany P, Steel EA, Avery K, Yoder N, Busack C, Thompson B (2010) Dealing with uncertainty in ecosystem models: lessons from a complex salmon model. Ecol Appl 20:465-482

Nash, J. F. 1951. Non-cooperative games. Ann.Math., 54: 286–295.

Niemeijer, David, and Rudolf S. de Groot. "A conceptual framework for selecting environmental indicator sets." *Ecological indicators* 8.1 (2008): 14-25.

Pauly et al., 1998. Fishing-down marine food webs. Science 279: 860-863.

Pauly, Daniel, and Reg Watson. "Background and interpretation of the 'Marine Trophic Index'as a measure of biodiversity." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360.1454 (2005): 415-423.

Plagányi, Éva E. *Models for an ecosystem approach to fisheries*. No. 477. Food & Agriculture Org., 2007.

Plaganyi et al., 2014. Multispecies fisheries managemet and conservation: tactical applications using models of intermediate complexity. Fish and Fisheries 15: 1-22.

Pope, J. G. 1989a. Simple representations of the yield surface.Working paper for the 1989 meeting of the ICES Multispecies Assessment Working Group. 4 pp

Pope, J. G. 1989b. Going straight for the Jacobian; Jonessan? Working paper to the 1989 meeting of the ICES Multispecies Assessment Working Group. 4 pp available as annex B on WGSAMs site.

Pope, J. G., Rice, J. C., Daan, N., Jennings, S., and Gislason, H. 2006. Modelling an exploited marine fish community with 15 parameters - results from a simple size-based model. -ICES Journal of Marine Science, 63: 1029-1044.

Rice, J. C ., D aan, N ., Pope, J. G ., and Gislason, H. 1991. T he stability of estim ates of suitabilities in M SVPA over four years o f data from predator stomachs. - ICES mar. Sei. Symp., 193: 34-45.

Rice, Jake C., and Marie-Joëlle Rochet. "A framework for selecting a suite of indicators for fisheries management." *ICES Journal of Marine Science: Journal du Conseil* 62.3 (2005): 516-527.

Shannon, Claude E. *(July 1948).* "A Mathematical Theory of Communication". *Bell System Technical Journal. 27 (3): 379–423.* doi*:*10.1002/j.1538-7305.1948.tb01338.x*.*

Smith, A.D.M., Brown, C.J., Bulman, C.M., Fulton, E.A., Johnson, P., Kaplan, I.C., Lozano-Montes, H., Mackinson, S., Marzloff, M., Shannon, L.J., Shin, Y.-J. and Tam, J. 2011. Impacts of fishing low trophic level species on marine ecosystems. *Science*, 333: 1147-1150

Stefansson G., Palsson O.K. 1997. Bormicon. A Boreal Migration and Consumption Model. Marine Research Institute, Rep. 58.

Stefansson G. 2001. Multi-species and ecosystem models. Reykjavik conference on responsible fisheries in the marine ecosystem. Reykjavik, Iceland 1-4 October 2001.

Thorpe, R.B., W.J.F. Le Quesne, F. Luxford, J.S. Collie and S. Jennings. 2015. Evaluation and management implications of uncertainty in a multispecies size-structured model of population and community responses to fishing. Methods in Ecology and Evolution 2015, 6, 49–58, doi: 10.1111/2041–210X.12292

Thorpe, R. B., Dolder, P. J., Reeves, S., Robinson, P., and Jennings, S. 2016. Assessing fishery and ecological consequences of alternate management options for multispecies fisheries. – ICES Journal of Marine Science, doi: 10.1093/icesjms/fsw028.